Egocentric Interaction for Ambient Intelligence

Dipak Surie


Department of Computing Science
Umeå University
SE-901 87 Umeå
Sweden
Abstract

Ambient intelligence refers to the vision of computationally augmented everyday environments that are sensitive, adaptive and responsive to humans and intelligently support their daily lives. Ambient ecologies are the infrastructures of ambient intelligence. To enable system developers to frame and manage the dynamic and complex interaction of humans with ambient ecologies consisting of a mixture of physical (real) and virtual (digital) objects, novel interaction paradigms are needed.

Traditional interaction paradigms like the WIMP (windows, icon, menus, and pointing devices) paradigm for desktop computing operate in a closed world, unaware of the physical, social and cultural context. They restrict human perception and action capabilities to screen, mouse and keyboard with the assumption that human attention will be fully devoted to interaction with the computer. Emerging interaction paradigms for ambient intelligence are typically centered on specific devices, specific computing environments or specific human capabilities. Also, many of them are driven by technological advancements rather than taking the human agent as their starting point. A principled, theoretical approach centered in the individual human agent, their situation and activities that are comprehensive and integrated while at the same time instrumental in the design of ambient ecologies has been lacking.

This thesis introduces egocentric interaction as an approach towards the modeling of ambient ecologies with the distinguishing feature of taking the human agent’s body, situation and activities as center of reference, as opposed to the more common device-centric approaches in facilitating human-environment interaction. Egocentric interaction is encapsulated in a number of assumptions and principles such as situatedness, the proximity principle, the physical-virtual equity principle, perception and action instead of “input” and “output,” and activity-centeredness. A situative space model is proposed based on some of these principles. It is intended to capture what a specific human agent can perceive and not perceive, reach and not reach at any given moment in time. The situative space model is for the egocentric interaction paradigm what the virtual desktop is for the WIMP interaction paradigm: more or less everything of interest to a specific human agent is assumed and supposed to happen here.

In addition, the conception and implementation of the easy ADL ecology based on egocentric interaction, comprising of smart objects, a personal activity-centric middleware, ambient intelligence applications aimed at everyday activity support, and a human agent literally in the middle of it all is described. The middleware was developed to address four important challenges in ambient intelligence: (1) tracking and managing smart objects and their state changes; (2) tracking a human agent’s situative spaces in terms of available physical and virtual objects; (3) recognizing human activities and actions based on interaction with physical and virtual objects; (4) managing and facilitating egocentric interaction with ambient intelligence.
applications; and to ease up the development of ambient intelligence applications.

The easy ADL ecology was first simulated in immersive virtual reality, and then set up physically as a living laboratory to evaluate: (1) the technological and technical performance of individual middleware components, (2) to perform a user experience evaluation assessing various aspects of user satisfaction in relation to the support offered by the easy ADL ecology, and (3) to use it as a research test bed for addressing challenges in ambient intelligence. While it is problematic to directly compare the “proof-of-concept” easy ADL ecology with related research efforts, it is clear from the user experience evaluation that the subjects were positive with the services it offered.
Dedication

“To Thomas and Lars-Erik for painting novel ideas into a beautiful picture”
Scenario

Before reading this thesis, the readers are requested to go through the scenarios about Gunnar Sjögren from a first person’s perspective. The scenario in “Getting into Gunnar’s shoes” is intended for introducing one potential application area where the work to be presented in this thesis fits in and the scenario in “Gunnar steps into the easy ADL ecology” is used as science fiction scenarios that motivate and guide further research in the field of Ambient Intelligence (Aarts and Grotenhuis, 2009, Aarts and Ruyter, 2009, Aarts and Wichert, 2009, Aarts and Encarnação, 2008). The scenario is also useful for introducing this thesis work to a wide range of audience with varying knowledge in the field of ambient intelligence.

Getting into Gunnar’s shoes

I like to browse through my photo album with a cup of coffee when I feel lonely. This photo was taken during my working days as a carpenter. I had a head injury during a construction work in Lycksele in 1992. The injury has affected me psychologically leading to small memory problems throughout my life. I am Gunnar Sjögren, a 64 year old pensioner living in Umeå, Sweden.

This is not my favorite picture, my stuga (cottage) in Ammarnäs half burnt. I forgot to turn off the stove and went to take a sauna. When I came back, all that I saw was my home burning like the wooden logs in a sauna. This picture reminds me that I need to turn off the stove after cooking. Ting ting, the door bell is ringing... the postman! I have got a new post from Norrlands University hospital. I have got an appointment on the 26th for my fortnightly checkucliparp. I will be out of town on this date, and should probably call them tomorrow, if I remember! I am currently under medication, but I often forget to take my medicines on time. That’s when the whole world spins on my head.
The situation is never better when I forget if I had taken my medicines or not. I panic and call my care-giver. Linda is nice, but I feel embarrassed to call her for such simple reasons often.

Look here, this photo was taken during the millennium New Year celebrations. Yes, I look pretty cold. It was a freezing evening and I forgot my gloves, cap and scarf at home. After the celebrations, I rushed back to my home and began to fiddle around with my apartment key using my frozen hands to unlock the main door, and you know what, I ended up disturbing Linda’s New Year party to open my apartment’s main door. I am not in full control inside my home either. I often forget where my toilet is during night times and land up in the kitchen. Then, I usually try calling Linda and since my phone buttons are too small, I end up calling wrong numbers. I often wonder why can’t there be a system that can guide me to the toilet at nights. I need more coffee before I can continue. “Hej coffee machine, can you prepare more coffee?” Nej! I am searching for kanelbullar (Swedish cinnamon roll) that I prepared with Ingrid, a friend of mine last time around. Here it is, but looks old and so are many other food items here. I should have thrown them away long time ago!

My phone is ringing, it is Erik my son, wait a minute. He is busy as usual, wants me to prepare the shopping list so that we could go to the supermarket this evening. Preparing a shopping list is not the easiest thing in the world. I don’t want to buy the whole supermarket, but don’t want to miss something important either. Damn, I have spilled some coffee on my shirt. Now, I remember, there are many other clothes that are to be washed. But why have I not washed them? Yeah, I was out of washing powder last time around, should probably include it in the shopping list. I am getting confused, what else to include in the shopping list? Milk, shower cream, tooth
pasta, veggies ... Oj då, the shower handle has fallen on the floor, and there is so much water in the bathroom. That’s why; the tap seems to be open. I am usually cautious in these circumstances since I have a history of falling down in the bathroom. Injuries, frustration, phone calls, hospital... I want to avoid it!

That’s a gift from Ingrid. I have not opened it for a week now, just want to keep it secretly for when I open it, my imagination might fade. We take dinner together at times, and usually I want to be the chef with some surprise menu. But, I find it hard to decide on the food menu based on the ingredients that I have at home, and these old recipe books. Mind you, I want to prepare a good dinner, but then who can guide me through with some interesting recipes. Ingrid checks my apartment and tells me what I could potentially prepare, but that’s not interesting from my perspective. Also, it would be nice to have perfect lighting and music to go with some good food.

I like to be social, but then it gets really difficult to move out of my home to meet friends. I used to discuss politics for hours with my friends at the biblioteket cafeteria when I was young and healthy. I can check if any of my friends are available online. I like what this computer does, but then it is tricky to use it at times. Paying my monthly bills using internet often puts me in a nightmare. So does checking horse racing updates, playing online memory games, and reading books. I prefer to read coffee table books or play physical memory games instead of using a computer for entertainment, but then computers contain unlimited games, books, movies... That’s my granddaughter Jenny in the wallpaper background. I tell her night time stories using this computer. I have a webcam, but it more often doesn’t work. Erik has tried hard to
teach me how to configure my webcam, but my expertise is not with computers. I can manage vacuum cleaners, washing machines, coffee machines... but not computers.

That’s a chart that contains information about the activities that I should perform daily. Linda keeps track of them when she visits me daily. But when she is not there, I usually have a hard time completing some of those activities. Doctors say I have mild-dementia, but I do not believe it, or at least I don’t want to believe it. I have been an independent man throughout my life and would like to be so. I need to get going with other stuffs for the moment.

**Gunnar steps into the easy ADL ecology**

I like the coffee table, and of course this new coffee flavor from Gevalia. I purchased it in...hmm...just forgot! My coffee cup says ICA Maxi (supermarket), yeah it’s true. I moved into the easy ADL home about a year ago. I was part of Bostaden’s (housing company) smart home campaign and became lucky to get this apartment when the other applicants were reluctant to live in a smart home environment. The initial thought sounds geeky and even scary, but after a year in the easy ADL home, I do not have many regrets. This coffee table gives me all the news updates that I crave for, allows me to play memory games online by placing and moving around stuff on it, acts as a photo album that I can share with my online friends, and after all allows me to enjoy a nice fika (Swedish coffee break).

The informative art painting in front of me is a gift that I got from Erik for my Birthday last month. You see Erik climbing the Sognefjord along with Jenny in Norway last summer. They are about to reach the summit, which is actually an indication that my kanelbullar in the oven is about to be ready soon. I can use the same painting frame for other information needs like knowing the next bus to the city center, discuss politics with my online friends, get information about everyday activities that I have forgotten, etc. It feels like a well thought out gift for me. Let me ask my coffee machine to prepare some coffee, love it with a kanelbullar. Ursäkta, need to visit the toilet now. You see the floor carpet that creates a virtual flower bed along the path that I should
take to reach the toilet. I don't have to bother Linda anymore to know my way around in my apartment. She gets an update of my activities in her mobile phone, so she can actually monitor my wellbeing from a distance. I don't want to be a part of the big brother camera, but then Linda monitors only those everyday activities that I need assistance. I like to do things in private after all, shhhhh...

The bathroom mirror begins to present my evening schedule. Shopping with Erik, this is actually in 30 minutes. The bathroom mirror asks me if I need some help from the shopping assistant. Ja! and I begin to have a quick shave. The shopping assistant checks for the items that are available in the easy ADL home based on predicting my potential activities in the coming week. The shopping assistant is smart; it keeps track of the current offers and includes those additional items that are worth buying into the shopping list, of course with my concern. Gone are those days when I get stressed to prepare a shopping list before Erik takes me to the supermarket. BTW, the shopping list usually appears in my shopping cart at the supermarket, so no worries of misplacing the shopping list somewhere in the home or forgetting to bring the shopping list to the supermarket.

The bathroom mirror begins to display fire flames, and my razor gets turned off. Oj oj, the kanelbullar is ready and I should take it out immediately or else it will get burnt. The stove display presents the shopping list and informs me about the lobsters on sale. Without any hesitation, I add it to the shopping list by moving the lobster icon to the shopping cart icon on the stove display using hand gestures. The stove is smart: it knows that my hand gestures are not for turning on the oven once again.

The medicine cabinet is trying to speak to me, maybe there is something important. I've got a message from the hospital to confirm my fortnightly checkup. The good thing is that I have the option to see all the dates and the times that I could potentially visit them in the next couple of weeks. Also, if I would prefer a checkup at home, such an option is available as well. I continue, “22nd at 10 o'clock in the morning should be fine” and the medicine cabinet gives me a confirmation.

After a quick shave, I begin to get ready. What should I wear? The wardrobe begins to present information about the expected weather conditions in the evening. It’s going to be really cold, -18°C with 50% chance of a snow storm. The wardrobe suggests several clothing options depending on the clothes that are already washed, and the expected weather conditions. I
will go with the second option. The wardrobe’s robotic arm gives me the clothes and helps me to wear them. Trevlig... looks good in this outfit!

It's a Friday evening and I met Ingrid to my surprise while shopping. We decided to have a calm evening at my place, and not to mention a sauna before. I can access my home sauna through my walking stick from a distance. Just need to tap the walking stick a few times on the ground and speak out “turn on the sauna”. It all looks science fiction, but after all I don't mind using technology if it is simple, useful and guess what I don’t have to wait for 20 minutes before the sauna gets red hot ready. I am on a pension and do not want expensive electricity bill every month. The easy ADL home understands my financial concern and turns off the apartment heaters, lighting, music, TV, etc. and waits until I get back home to turn them on. Smart is my easy ADL home; it turns on the heaters alone when I am at a close proximity to my home. Heating up my place takes a few minutes, if not more at times and I don't want to put myself and Ingrid in cold anyway.

I arrived home, and my hands are already frozen. The easy ADL home recognizes my smart bracelet and opens the main door. I get help for placing the shopped items in the right place. Frozen items should not be stored in the fridge; toiletries should not be in the same rack as food items, etc. Ingrid arrives home. The door welcomes her and the home profile is automatically set to “Gunnar med Ingrid”. After a relaxing sauna, we reached the cozy living room couch. Candle lights began to turn on, and its aromatic smell marked the tone for the evening after the sauna. The background music is one of Ingrid’s favorite and it’s time for deciding the dinner menu. The recipe assistant checks the apartment for available raw materials for preparing a special dinner. It then checks with Ingrid’s food profile and my cooking skills. The recipe assistant makes a smart decision based on the above mentioned factors to present the set of potential dinner menu options on the living room wall display. “Let’s go for some Lobster Thermidor”, I said looking into Ingrid’s eyes. She seems to be excited and replied absolu! I went to the kitchen to surprise her with a glass of red wine. My fridge is smart, it knows that red wine doesn’t go well with Lobster Thermidor, and gently suggests me to go for Bordeaux’ Graves or Rhone’s Hermitage (white wine). Here you go some white wine. Ingrid looked at me and said, “Great! It’s not the red one”.
I asked Ingrid to not come to the kitchen so that the chef can give a pleasant surprise. Ingrid smiled and moved towards the bookshelf and grabbed a new night time story book. The book connected Ingrid to Jenny, my granddaughter who was online waiting for me to tell her bedtime stories. Jenny knows Ingrid who took my role, while I was informed about it in the kitchen through the smart sauce pan. Ingrid decides to give Jenny a realistic story experience by associating the objects on the living room table to characters part of the story and begins to manipulate them. Jenny could not only hear Ingrid, but also experience the story enacted virtually through cartoon characters on her bedroom pillow, blanket and the ceiling. As Jenny begins to sleep, Ingrid holding an object associated with a cartoon character begins to vibrate with a snoring sound. Ingrid makes sure that Jenny is sleeping, and taps her couch to recline it and to get some massage while enjoying Rhone’s Hermitage. The smart couch takes care of the home’s special guest.

In the mean while, I pre-prepared lobster meat and placed them back into the shell. Suddenly, I am confused in preparing the sauce. Should I add lemon juice? Oj... I don’t know. The fridge understands that I am confused, and by knowing that I am preparing the sauce presents the sauce recipe. Tack! I took a lump of butter on a pan, added shallots and cooked it for a while. I moved on to put the prepared sauce over the lobster to grill it for a few minutes. The baking dish begins to project information. The sauce is not properly prepared, check this recipe. Nej men... is there still a problem with this sauce? Oj, I have forgotten to add mustard, herbs and seasoning. I was about to make a ‘wrong’ster thermidor!

Golden brown lobsters are ready and it’s time to setup the dinner table. Where are the knives and the forks? Not there in my kitchen drawer! Has someone stolen it from my home? But then why just knives and forks... I am getting paranoid and need them to
setup the table! Should I call Linda? But then I am not disturbing her these days. The dish washer is trying to say something to me, but I am too busy mate! Why are the cotton napkins changing color? It is trying to say something... okej, it’s in the dishwasher.

Ingrid is invited to the dining hall, and after tasting a piece of lobster, she said, “Javisst...that’s delicious”. The dinner was pleasant until the medicine cabinet got upset with me for not taking my medication. The medicine cabinet selects the right medicines for me, and drops them on the tray. I took the medicines, and the tray acknowledged it with a smile. Off we went to sleep. Get back tomorrow, till then... sov gott.
Acknowledgements

“At times our own light goes out and is rekindled by a spark from another person. Each of us has cause to think with deep gratitude of those who have lighted the flame within us.” — Albert Schweitzer

Thomas (Pederson) for introducing me to the field of Ubiquitous Computing, for being a creative thinker willing to view things from unconventional directions, for being a compassionate leader, friend and a nice human being. Lars-Erik Janlert for opening the doors to Cognitive Science, and for being a rainbow in the clouds of open-mindedness and practicality.

Fabien, Daniel (Sjölie), Olivier, Florian, Dilip, Thomas (Johansson), Anders (Backman), Kenneth (Bodin), Gösta, Björn (Sondell), Stefan, Samuel, Erik (Lövborg), Robert (Bhatt) and other easy ADL project members for the valuable discussions, system development and fun moments. The easyADL project was funded from the EC Target 1 structural fund program for Northern Norrland, Sweden.

Walls are meant not to keep people out but to see who cares enough to break them down like Annabella, Erik (Billing), Ola, Johan (Tordsson), Helena (Lindgren), P-O., Claude, Mostafa, Farahnaz, Mina, Johanna & Henrik, and Thomas (Nilsson). Appreciate it. Mats, Bertil, Inger, Yvonne, Anne-Lie, Tomas (Forsman), and Per (Lindström) for your support activities. Erik (Elmroth), Lennart, Frank, Bo, Patrik, Jürgen, Michael (Minock), Francisco, Anders (Broberg), Mikael (Hansson), and Häkan (Gullikson) for being kind, concerned and guiding my PhD studies at moments in time.

Great friends are hard to find, difficult to leave, and impossible to forget: Arsalan, Kairul, Mukta, Zohreh, Lokesh, Sidd, Bhaskar, Srinivasa, Kumar, Suman, Murali, Venky, Amman, Shiplu, Shafiq, Sampath, Raghuraman, Akil, Edward, Rameez, Mahesh (Nalluri), Nippu, Azzizi, Uday, Anubhav, Manna, Muninder & Sherwani, Kiran & Sireesha, Ravi & Vidya, Bala & Uma, Tony & Jothi, Martin & Lyna, German, Masoud & Julia, Andinet & Jerry, Yared & Hanna, Michael, Bruno, Sundip, Steffi, Huo, Kamal (RMK), Shilpa, Karthik (TP road), Kaushik (Valluvar kottam), Arun, Veronica, Gunn & Sven, Björn, Bernt-Erik, Noori, and Diana.

Would like to thank you from the bottom of my heart, but I wonder if my heart has a bottom for you guys in my life: Subadra, Suresh, Aish, Atti, Veda, Ruku, Lakshmi & Co.,Pachu, and other family members. I can repay anything material in this universe, but would die forever in debt for your unconditional love towards me: thanks Rathi. Viysh, the cutiepie of my life. Is it time to join Cramer in search for an adjoint matrix?
# Contents

## Introduction

1.1 From Science Fiction to Reality  
1.2 Mainframe, Desktop and Mobile Computing  
1.3 Ubiquitous Computing  
1.4 Today’s Everyday Artifacts  
1.5 Spot Light on Human Agents  
1.6 Ambient Intelligence  
1.7 Scope and Objectives  
1.7.1 Explore a Novel Interaction Paradigm  
1.7.2 Develop an Infrastructural Platform  
1.8 Research Methodology  
1.9 Contributions  
1.10 Thesis Outline  
1.11 Publications

## Emerging Interaction Paradigms and Approaches

2.1 Traditional Interaction Paradigms  
2.1.1 Command-Line and Textual Interfaces  
2.1.2 Graphical User Interfaces  
2.2 Emerging Interaction Paradigms  
2.2.1 Tangible User Interfaces  
2.2.2 Surface Computing  
2.2.3 Ambient Displays and Informative Art  
2.2.4 Multimodal Interaction  
2.2.5 Affective User Interfaces  
2.2.6 Perceptual User Interfaces  
2.2.7 Attentive User Interfaces  
2.2.8 Context-Aware Computing  
2.2.9 Implicit Human-Computer Interaction  
2.2.10 Activity-Based Computing  
2.2.11 Embodied Interaction
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2.12</td>
<td>Dimension Space Model</td>
<td>54</td>
</tr>
<tr>
<td>2.2.13</td>
<td>Reality-Based Interaction</td>
<td>54</td>
</tr>
<tr>
<td>2.2.14</td>
<td>Mixed-Initiative Interaction</td>
<td>55</td>
</tr>
<tr>
<td>2.2.15</td>
<td>Human-Centered Computing</td>
<td>56</td>
</tr>
<tr>
<td>2.2.16</td>
<td>Discussion</td>
<td>56</td>
</tr>
<tr>
<td>2.3</td>
<td>Reality-Virtuality Continuum</td>
<td>57</td>
</tr>
<tr>
<td>2.3.1</td>
<td>RV Continuum and Ambient Intelligence</td>
<td>58</td>
</tr>
<tr>
<td>2.3.2</td>
<td>Augmented Reality and Mixed-Reality</td>
<td>59</td>
</tr>
<tr>
<td>2.4</td>
<td>Wearable Computing</td>
<td>62</td>
</tr>
<tr>
<td>2.4.1</td>
<td>Definitions of Wearable Computing</td>
<td>62</td>
</tr>
<tr>
<td>2.4.2</td>
<td>Wearable Computing as an Infrastructure for AmI</td>
<td>62</td>
</tr>
<tr>
<td>2.4.3</td>
<td>Head-Mounted Display and QBIC Belt Computer</td>
<td>63</td>
</tr>
<tr>
<td>2.4.4</td>
<td>Gesture Pendant and “Sixth Sense” Computer</td>
<td>63</td>
</tr>
<tr>
<td>2.4.5</td>
<td>Electronic Textiles</td>
<td>64</td>
</tr>
<tr>
<td>2.4.6</td>
<td>Power Harnessing</td>
<td>65</td>
</tr>
<tr>
<td>2.4.7</td>
<td>Personal Server</td>
<td>65</td>
</tr>
<tr>
<td>2.4.8</td>
<td>Body Area Networks</td>
<td>66</td>
</tr>
<tr>
<td>2.4.9</td>
<td>Commercial Wearables</td>
<td>66</td>
</tr>
<tr>
<td>2.5</td>
<td>Smart Objects</td>
<td>67</td>
</tr>
<tr>
<td>2.5.1</td>
<td>Definitions of Smart Objects</td>
<td>68</td>
</tr>
<tr>
<td>2.5.2</td>
<td>Smart Objects as an Infrastructure for AmI</td>
<td>68</td>
</tr>
<tr>
<td>2.5.3</td>
<td>Media Cup</td>
<td>69</td>
</tr>
<tr>
<td>2.5.4</td>
<td>Sentient Artifacts and AwareMirror</td>
<td>69</td>
</tr>
<tr>
<td>2.5.5</td>
<td>Co-operative Artefacts</td>
<td>70</td>
</tr>
<tr>
<td>2.5.6</td>
<td>Smart Pizza Packing</td>
<td>71</td>
</tr>
<tr>
<td>2.5.7</td>
<td>Smart Floor</td>
<td>72</td>
</tr>
<tr>
<td>2.5.8</td>
<td>booTable</td>
<td>72</td>
</tr>
<tr>
<td>2.5.9</td>
<td>Commercial Smart Objects</td>
<td>73</td>
</tr>
<tr>
<td>2.5.10</td>
<td>Interaction with Smart Objects</td>
<td>73</td>
</tr>
<tr>
<td>2.6</td>
<td>Smart Environments</td>
<td>75</td>
</tr>
<tr>
<td>2.6.1</td>
<td>Definitions of Smart Environments</td>
<td>76</td>
</tr>
<tr>
<td>2.6.2</td>
<td>Smart Environments as Infrastructure for AmI</td>
<td>77</td>
</tr>
<tr>
<td>2.6.3</td>
<td>Smart Homes</td>
<td>77</td>
</tr>
<tr>
<td>2.6.4</td>
<td>MavHome and Neural Network House</td>
<td>77</td>
</tr>
</tbody>
</table>
2.6.5 iDorm
2.6.6 Gator Tech Smart House
2.6.7 Aware Home
2.6.8 Nursebot Project, SmartBo, Gloucester House and CareLab
2.6.9 Typical Challenges within Smart Homes

Egocentric Interaction: a Human-Centered Interaction Paradigm

3.1 Understanding Human Cognition
3.1.1 Theories Related to Human Cognition
3.1.2 Cognitive Architectures
3.1.3 Perception and Action
3.1.4 Attention and Intention

3.2 Basic Principles and Assumptions
3.2.1 Terminological Shift from “User” to “Agent”
3.2.2 Merging Virtuality, Mobility and Ubiquity Paradigms
3.2.3 Different from Classical User-Centered Approaches
3.2.4 Situatedness: The View from Here
3.2.5 The Proximity Principle
3.2.6 Perception-Action instead of Input-Output
3.2.7 The Physical-Virtual Equity Principle
3.2.8 Summary of Basic Principles and Assumptions

3.3 A Physical-Virtual Design Perspective
3.3.1 Physical-Virtual Environments
3.3.2 Physical-Virtual Artefacts
3.3.3 Physical-Virtual Situations
3.3.4 Physical-Virtual Activities

3.4 Situative Space Model
3.4.1 Main Components of the Model

3.5 Applying the Situative Space Model
3.5.1 Situation

3.6 Associating the SSM to Human Intention and Attention

3.7 Discussion
The easy ADL Ecology: an Infrastructure for Ambient Intelligence

4.1 Introduction 123

4.1.1 Ambient Ecology 124

4.2 The easy ADL Ecology 125

4.2.1 Immersive Virtual Reality Home Environment 125

4.2.2 Living Laboratory Physical Home Environment 127

4.2.3 Infrastructure of the easy ADL Ecology: Overview 127

4.3 Smart Objects 128

4.3.1 Heterogeneity of Smart Objects 129

4.3.2 Entity-Relationship Representation of Smart Objects 135

4.4 A Personal Activity-Centric Middleware 137

4.4.1 Object Manager 138

4.4.2 Situation Monitor 140

4.4.3 Activity Recognizer 140

4.4.4 Interaction Manager 141

4.5 Ambient Intelligence Applications 142

4.6 Interaction Management Rules 143

4.7 Interaction Techniques 147

4.7.1 Gesture based Object Manipulation 149

4.7.2 Speech based Object Manipulation 149

4.7.3 Distributed Visual Presentation of Objects 150

4.7.4 Time-Multiplexed Audial Presentation of Object 150

User Experience Evaluation of the easy ADL Ecology 153

5.1 Introduction 153

5.2 Evaluation Setup 155

5.2.1 The Living Laboratory Home Environment 155

5.2.2 Participants 155

5.2.3 Scenarios within the easy ADL Ecology 156

5.2.4 Wizard-of-Oz Method 156

5.3 Methodology 158

5.3.1 Establishing the UX Factors 158
5.3.2 UX is Momentary, Long-term and Anticipatory Experience 158
5.3.3 Within and Without the easy ADL Ecology 159
5.3.4 UX Scores and their Value 159

5.4 Results and Discussion 160
5.4.1 Human Physical-Virtual Activities 160
5.4.2 Human Situatedness 162
5.4.3 Human-Centered Load during Interaction 163
5.4.4 Human Attention 163
5.4.5 Multimodal Interaction with Virtual Objects 164
5.4.6 Social Context 164
5.4.7 Design of Smart Objects 165
5.4.8 Additional Remarks by the Subjects 166

Tracking State Changes of Physical and Virtual Objects 168
6.1 Introduction 168
6.2 Wireless Sensor Networking of Smart Objects 169
6.2.1 What are Wireless Sensor Networks? 169
6.2.2 General Requirements 170
6.2.3 System Description 170
6.2.4 Experimental Setup 175
6.2.5 Evaluation 177
6.3 Tracking State Changes to Virtual Objects 180

Situative Space Tracking across Multiple Modalities 183
7.1 Introduction 183
7.1.1 SSM across Multiple Modalities 184
7.1.2 Establishing the Situative Space Boundaries 184
7.2 Approaches to Situative Space Tracking 185
7.2.1 Discussion 188
7.3 Situative Space Tracking System I 189
7.3.1 System Architecture 190
7.3.2 Evaluation I 195
Chapter 1

Introduction

This chapter introduces ambient intelligence and provides an overview of the thesis. A brief history of the evolution of computing, and insight into everyday artifacts and their role in everyday environments are discussed. The aims, scope, research methodology and the contributions made are briefed. Also, the thesis alignment into chapters and relevant publications are presented.

1.1 From Science Fiction to Reality

The scenarios where in “Gunnar steps into the easy ADL ecology” might well be considered a science fiction even 20 years ago, but today with the advancements in processing, memory, storage, communication, networking, sensing, actuation, interface, and interaction technologies the stage is set for making it a reality in the future. This thesis will attempt to play its part in moving towards such a reality in the future with contributions in the field of Ambient Intelligence (Aarts and Grotenhuis, 2009, Aarts and Ruyter, 2009, Aarts and Wichert, 2009, Aarts and Encarnação, 2008). Other terms like Ubiquitous Computing (Weiser, 1995, Bell and Dourish, 2007), Pervasive Computing (Satyanarayanan, 2001), Smart Environments (Cook and Das, 2007, Cook and Das, 2004), Wearable Computing (Starner, 2001, Mann, 1997), Smart Objects (Beigl and Gellersen, 2003, Kawsar et al., 2005), Augmented Reality (Mackay et al., 1998) or Mixed Reality (Costanza et al., 2009), Calm Technology (Weiser and Brown, 1997, Weiser and Brown, 1996), Invisible Computer (Norman, 1998b), Disappearing Computer (Streitz et al., 2007, Russell et al., 2005), Everyware (Greenfield, 2006), Ambient Ecology (Goumopoulos and Kameas, 2009, Kameas et al., 2009), etc. are comparable and closely related to the term AmI in the sense that they all share similar visions towards a post-desktop model of computing and Human-Computer Interaction (HCI) that takes place in the everyday physical world, but with minor differences in their focus. Some of the above mentioned visionary terms are discussed further in different parts of this thesis.

1.2 Mainframe, Desktop and Mobile Computing

Taking a break from the visionary terms and getting back in time to the history of computers, mainframe computers were the dominant first generation computers manufactured in the late 1950s. They were big, expensive and used only by experts. Large companies were interested in mainframe computers for their ability to calculate as a replacement for
calculators and it explains the reason for their poor interaction capabilities with human agents: human computer interaction was not the focus of mainframe computers. Punched cards, magnetic tapes, etc. were used to provide input for batch processing and many people shared a single mainframe computer. Refer to Fig. 1.1.

The second generation computers are mainly desktop computers where a human agent interacts with one computer at a time. Today, desktop computers are probably the most popular and well known form of computers that operates based on the Windows, Icons, Menus and Pointing Devices (WIMP) interaction paradigm. Human agents perform personal activities (usually virtual activities (Pederson, 2003)) using desktop computers that usually comprises of a monitor, a keyboard and a mouse. As long as human agents are satisfied with desktop computers that are well established as computers, what are the needs to think beyond desktop computing? Desktop computers will probably be more stable and eliminate some of the additional challenges in building computers based on above mentioned visionary terms, at least in the near future. In fact desktop computers remain silent in a corner of an environment and do not bother to invade a human agent’s physical environment or even their life in the pretext of providing computing support. They are still useful in providing computing support for human virtual activities like paying bills through the internet or exchanging emails or reading news online. Is it not better to use the well-established desktop computers instead of going for more complicated visions like ambient intelligence? This is more a philosophical question and this thesis will not focus on why we are doing what we are doing - research in ambient intelligence. However, this thesis will continue to argue that: (1) there are everyday situations in which ambient intelligence can potentially be beneficial to a human agent than just to bother them (refer to Gunnar’s scenarios); and (2) with the advancement in technology we are closer in achieving the visions of ambient intelligence.

One of the reasons why desktop computers are popular is because they brought personal computing to the foreground at an affordable cost. Desktop computers are based on an office metaphor or to be more precise on a desktop metaphor. Is it possible to replace a physical desk in an office with folders, files, pencils, erasers, etc. within a magical box like machine that can do computation and facilitate office virtual activities? Desktop computers have answered this question rather successfully. However, the interaction capability of a desktop computer is limited to an office setup and its affiliated activities. It was meant for white collar workers in an office and was not developed primarily for blue collar workers, disabled and elderly people, children, athletes, or saints! Of course with little tuning here-and-there, it is possible to fit more user groups but the primary purpose of it as a general purpose machine with a desktop metaphor suitable for performing virtual office activities will not change. According to (Norman, 1998b), the advantage of desktop computers being all-in-one general purpose machine fade with the fact that desktop computers makes it complex for a human agent to use all its functionalities on offer introducing usability issues. Also, its general purpose nature restricts the user groups and the contexts in which its computing capabilities could be used. In contrast, individual human beings are unique in
their own ways with diverse capabilities and skills in the physical world in comparison to the interaction possibilities offered by desktop computers.

Human agents are mobile agents and the form factor of desktop computers makes it difficult to be used in mobile context. With the miniaturization of computer parts, we are blessed with mobile computers (like smart phones, PDAs, etc.) that are powerful enough to satisfy a human agent’s mobile computing needs. Research efforts in mobile user interface design have enabled human agents to interact with computers on the go. However, apart from the adaptation to a smaller screen, and touch-based gesture support, mobile user-interfaces have not made justice to the wealth of technological advancements that could potentially enable mobile computers to be used in varied mobile context. Human agent’s interaction with mobile computers has not changed a lot (they use a polished version of the same old WIMP interaction paradigm). It is still difficult to read an online newspaper or send an email while bicycling!

Beyond technological factors, other factors especially human skills, capabilities and limitations play an important part in facilitating HCI. Desktop computers assume that human agents can devote their complete attention while interacting with them, and that there is nothing else that is important enough to attract a human agent’s attention in the physical world. Such an assumption about human agents restricts the context in which computers could be used. We are at a juncture where the technological advancements allow us to either polish the existing interaction paradigm (here the WIMP interaction paradigm is referred) or to re-think and develop novel interaction paradigms that would handle a human agent’s attention capabilities better, similar to the one to be described in this thesis (egocentric interaction).

Similar arguments could be developed of the input and the output capabilities offered by desktop computers (including mobile computers) and their restriction of computer use to specific context. Human perceptual capabilities are beyond visual modality including audial and tactile modalities, and more often human agents use multiple modalities to perceive information from their environment. By having visual monitors of typically similar embodiment and dimensions, and often just one monitor per computer, desktop computers restrict the space through which a human agent can perceive virtual information (output information from computing systems).

Human action capabilities are beyond keyboard and mouse manipulations typical with desktop computers. Human agents can grab, hold, manipulate, arrange and store physical objects. Such skills are probably the basic skills that human agents learn when they are still a toddler. Human agents communicate naturally and explicitly with other human agents through multiple modalities including speech, gestures, touch, etc. Often, they also exchange information that is not explicitly conveyed, but rather implicitly through their emotions, operations, actions and the context in which they are situated. Desktop computer restricts a human agent’s action space in which they could perform virtual actions and provide information (input) to computing systems.
To summarize, desktop computers operate in a closed world and are not aware of a human agent’s contextual information in the physical world. They support personal virtual activities, but ignore the fact that human agents do have physical, social and cultural context and that they might prefer to perform physical activities in parallel to performing virtual activities using computing systems. Also, human agents might need support for their physical activities which is rather complicated (if at all possible) using desktop computers that are not context-aware (Abowd et al., 1999) and are not able to intelligently adapt their behavior dynamically. The above mentioned limitations motivate research efforts to explore computing possibilities beyond the well-known desktop computing.

**Fig.1.1.** Three phases of computer user namely mainframe computing, desktop computing and ubiquitous computing (Weiser, 1995).

### 1.3 Ubiquitous Computing

Late Mark Weiser often referred to as the father of ubiquitous computing from Xerox Palo Alto Research Center coined the term *ubiquitous computing* (Weiser, 1995). According to him, ubiquitous computing marked the third wave in computing following the era of mainframe computing and desktop computing, where a human agent interacts with many computers at the same time (refer to Fig. 1.1). The computing technology is expected to be calm and reside in the background of everyday human life, but still ubiquitously available and play a central role in supporting everyday activities. His famous quote, “The most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it” refers to the disappearance and the seamless integration of any technology that is successful into the realm of everyday life. Such profound technologies fit human environments and their lives naturally, and do not force human
agents to gain additional technological skills before making use of such technologies (Weiser, 1995). Typical examples include electric motors in a car and electric lamps in a building where the technology has diffused so much into our everyday life that we hardly speak of them as something special.

According to (Streitz, 2001), there are at least two ways in which computer technology could disappear into everyday environments occupied by human agents. One is physical disappearance due to the miniaturization of computer parts and the drastic reduction in their cost thereby enabling everyday artifacts like household objects, furniture, home appliances, wearable outfit, building infrastructure, etc. to be augmented with computing technology. The other is more a disappearance from a psychological perspective, i.e. human beings are so used to these computers embedded within everyday artifacts that they do not perceive them as computers but instead view them as everyday artifacts that are useful for accomplishing their everyday activities. Even though artifacts like the living room wall display or the floor carpet described in Gunnar's scenario contain large computer parts that are physically perceivable by human agents, at least as interactive displays, they are still expected to be psychologically ignored as computers.

In Fig. 1.1, the era of ubiquitous computing is marked by multiple computers simultaneously interacting with a human agent in offering computing support. Such a perspective on computing support encourages further research on human-environment interaction where an environment contains multiple computing systems interacting with a human agent simultaneously. This thesis will address the challenges of designing and modeling human-environment interaction taking a human agent’s body as a center of reference referred by the term “egocentric interaction”. Human-centered factors like perception, action, attention and intention that often take a back seat in dealing with device-centric approaches surface to the top in addressing human-environment interaction.

Augmenting everyday environments with multiple computers affect the everyday setup in two ways: (1) the computing (or virtual) artifacts now becomes a part of the physical environment; and (2) the physical artifacts are sensed and actuated using computing technology, thereby becoming a part of the computing (or virtual) environment. Thus, ubiquitous computing environments allow for the co-existing of physical and virtual artifacts as described in mixed-reality research. This introduces the additional challenge of handling both physical and virtual objects when modeling human-environment interaction.

1.4 Today’s Everyday Artifacts

Stop reading this thesis and look around your environment to spot everyday (physical) artifacts that possess computing power. You would recognize that everyday artifacts like wrist watches, microwave ovens, fridges, washing machines, digital cameras, MP3 players, mobile phones, rice cookers, cars, alarm clocks, printers, baby gyms, etc. contain computing parts, usually sensors, actuators, microcontrollers, displays, and some buttons. But then why is it that computing technology has not pervaded our everyday
environments to the level envisioned by Mark Weiser? There are many reasons for it. To mention a few, today's everyday artifacts usually work locally and in isolation. They do not have the capabilities to co-operate and share their computing capabilities with other heterogeneous everyday artifacts or do not have the capabilities to make use of external surrogates (Satyanarayanan, 2001) like smart mobile phones or desktop computers. Everyday artifacts in isolation contribute less to the visions of ubiquitous computing in comparison to the holistic experience offered by a network of everyday artifacts populated in a human agent’s environment similar to the vision of the internet of things (Gershenfeld et al., 2004). Also, today's everyday artifacts are not connected with computing applications that make use of them as a resource for their operations. Does that mean that there are no killer applications that are interested in a network of everyday artifacts with computing power? The scenarios part of Gunnar's story could be one potential application area for the internet of things (Kortuem et al., 2010).

Everyday artifacts of today do not possess "intelligence". But then how does a rice cooker know when the rice is ready or an alarm clock that begins to ring on time or a toilet closet that automatically flushes after use! Probably they are just reacting to sensory or preprogrammed inputs instead of thinking, reasoning and adapting their behaviors according to the operational context which is usually considered as an indication of possessing artificial intelligence. Will a network of intelligent everyday artifacts take us to the vision of ubiquitous computing? Partially yes, but not completely because the everyday artifacts that we know of today have already established themselves as an inherent part of our environments, life and society. Human agents are aware of the role played by everyday artifacts in their life and are able to interact with them using a common understanding of the physical world. Augmenting computing and networking capabilities within everyday artifacts would not only introduce additional functionalities, but might also disrupt the simplicity of such everyday artifacts making it more challenging for a human agent to use it in everyday context (Norman, 1998a). Also, human agents interact with everyday artifacts in a natural manner making use of their skills, and capabilities. Augmented everyday artifacts might introduce interaction possibilities that are more complex and unnatural to human agents, there by just overwhelm them and defeat the visions of ubiquitous computing.

Everyday artifacts are not general purpose devices like desktop computers. It opens up the possibilities of augmenting individual everyday artifacts with computing and interaction capabilities that are diverse and unique there by allowing for the existence of a wide range of augmented everyday artifacts with varying computing resources suitable for specific setups and context. Some augmented everyday artifacts might simply present their identity information to the interested applications, while other augmented everyday artifacts might possess artificial software agents within it that allow for its existence independently. Note that such augmented everyday artifacts are referred to as smart objects in this thesis (Kawsar et al., 2008, Beigl and Gellersen, 2003).

Based on the above discussion there are 3 points that come to surface in exploring a novel interaction paradigm for ubiquitous computing: (1) the
environment containing future everyday artifacts should be networked to provide a holistic experience for a human agent instead of the services offered by individual everyday artifacts in isolation, thereby introducing a need to view the environment as an ecology of future everyday artifacts (or smart objects); (2) the environment containing future everyday artifacts should be intelligent and context-aware; and (3) the environment containing future everyday artifacts should be human-centered.

1.5 Spot Light on Human Agents

The technical and technological advancements like activity recognition, indoor location tracking, wireless sensor networks, context-aware middlewares, intelligent user-interfaces, etc. might be important and valuable from a ubiquitous computing research perspective. However, human agents who are the potential beneficiary of such technological advancements are not particularly keen on the technology itself, but instead on how such advancements would assist their activities in everyday settings. They care for the positive experience provided by environments augmented with technology in comparison to non-augmented everyday environments.

Everyday environments like homes, offices, airports, shopping malls, etc. are not machines for living within it. This makes everyday environments different to the desktop computing environments with aspects like physical, social and cultural context that are usually ignored for simplicity when dealing with computers suddenly pops-up as important issues to address. Human agents are not mere users of computer systems in an everyday environment. They have an everyday life going on with routines, practices, attitudes, moods, likes, dislikes, phobias, abilities, skills, limitations, priorities, preferences, relationships, etc. that cannot be ignored. They improvise their behavior according to context making it difficult to predict and plan for their future states. This introduces a need to model ubiquitous computing environments as dynamic, adaptive and personalized environments centered on human agents.

Egocentric interaction to be described in chapter 3 attempts to address some of the challenges concerning a human agent by taking a human-centered stance in proposing an interaction paradigm for ambient intelligence.

1.6 Ambient Intelligence

Ambient Intelligence (AmI) (Aarts and Encarnação, 2008, Aarts and Wichert, 2009) is built upon the ideas of ubiquitous computing, but is more specific in its human-centered vision for computing systems of the future. AmI refers to the vision of computationally augmented everyday environments that are sensitive, adaptive and responsive to the presence of human agents, and intelligently support them in their daily lives. Within the context of this thesis, the infrastructure for realizing ambient intelligence is referred to as ambient ecology.
The term *ambient* refers to surrounding space or environment, while the term *intelligence* refers to the ability to perceive, adapt, learn, reason and act in a situation. AmI envisions human agents to be in the centre of technological advancements, and attempts to improve their productivity, creativity, well-being and experience through enhanced and novel human–environment interaction. AmI emphasizes the importance of technology disappearing into the surrounding in providing computational support to human agents, similar to ubiquitous computing. In some sense, the difference between the terms ubiquitous computing and ambient intelligence might well be political with the term ubiquitous computing popular in North America while the term ambient intelligence is popular in Europe.

AmI envisions human environments to be populated with numerous, heterogeneous and interconnected computing devices with varying technological and computational capabilities that are often embedded within everyday objects, furniture, home appliances, building infrastructure, wearable outfit, etc. to operate collectively and create a sense of intelligence surrounding human agents. The aim is to consider the environment as a whole instead of viewing individual systems or devices in isolation. This shifts the challenge from human–computer interaction to human–environment interaction when dealing with ambient intelligence (as mentioned earlier).

AmI is a multidisciplinary research field including computer science, electronics and mechanical engineering, interaction design, industrial design, architecture and cognitive science. AmI is built upon several computing ideas including ubiquitous computing, artificial intelligence, context-aware computing, sensing and actuation, pervasive networking, and human-centered interaction and interfaces.

According to (Augusto, 2010), AmI system is defined by three essential elements namely environment, interaction constraints and interactors. An adapted version is shown in Fig. 1.2. The essential elements can further be refined depending on the application domain. The interaction constraints are an important element of AmI system since it facilitates interaction between the environment and the agents. This thesis will focus on the interaction constraints by proposing a novel interaction paradigm referred to as egocentric interaction with principles and assumptions, models, design perspectives, interaction management rules and techniques. Egocentric interaction, being more conceptual does not deal at the level of sensors, actuators, interfaces, etc. but considers a human agent’s body, mind, situation and activities as a starting point in facilitating interaction with the environment.

Getting closer towards the vision of AmI will change future everyday environments, people behaviors in those environments, and the way they perform their activities. Even though such a change at this moment is envisioned to be a positive one with potential benefits, there are many technical challenges like privacy, trust, security, etc. that will emerge. Ambient ecologies are likely to be aware of personal, social, behavioral, financial and medical data about human beings. Improper handling of such information might turn the landscape of ambient intelligence into a dark one. This thesis will however not dwell deeper into such challenges (not the focus), while such challenges are acknowledged.
1.7 Scope and Objectives

There are many problems and challenges in the field of ambient intelligence that span across technical, technological, design, personal, social, cultural, political, business and artistic domains. The research presented in this thesis will focus on the technical, technological and design aspects useful for designers and engineers in building ambient ecologies. Other aspects are important, but beyond the scope of this work. However, during the user experience evaluation of the easy ADL ecology described in chapter 5, the subjects were questioned about the personal, social and aesthetic aspects of being a part of the easy ADL ecology.

This work is specific to private environments like a home, or an office instead of public environments like an airport or a shopping mall. The proposed interaction paradigm or the underlying infrastructure will not change much if this work is to be ported to public environments, but additional challenges like privacy, trust, etc. will emerge and should be handled better. Even though, the issue of privacy and trust is addressed, it is at a level usually present within private environments.
The research work presented views an environment to be occupied by a single human agent. The concepts presented and the systems built are extendable to incorporate multiple human agents in an environment, but is left as future work for simplicity. The interaction paradigm and the infrastructure are intended to provide support for personal everyday activities while collaborative activities and activities that spans over a large period of time say in months or years are ignored, once again for simplicity.

1.7.1 Explore a Novel Interaction Paradigm

The primary objective is to explore an interaction paradigm suitable for ambient intelligence by viewing a human agent as a center of reference for modeling and facilitating interaction in their everyday setting doing activities in a natural manner. The aim is to develop models, principles and assumptions, design perspectives, interaction management rules and techniques that would help further understand the novel interaction paradigm and ease the development process.

A further goal is to provide: (1) an overview of the existing post-WIMP interaction paradigms; (2) to discuss their suitability for ambient intelligence and in facilitating human-environment interaction; and (3) to present egocentric interaction in context to such interaction paradigms and approaches.

1.7.2 Develop an Infrastructural Platform

The secondary objective is to develop an infrastructural platform for ambient intelligence (the easy ADL ecology) based on egocentric interaction. The easy ADL ecology comprises of smart objects, a middleware, a set of ambient intelligence applications and a human agent in the middle of it all.

Smart object related sub-goals include: (1) to provide an overview of existing approaches in designing and modeling smart objects and environments; (2) to present our approach in modeling smart objects and environments based on egocentric interaction.

Middleware related sub-goals include: (1) to explore the important and interesting challenges that a typical middleware should address in an ambient ecology; and (2) to design and build such a middleware that addresses those challenges. Existing middleware often focus on specific challenges which is important for obtaining sharp research results, but often are insufficient in an ambient ecology context where the big holistic picture should be the focus. The personal activity-centric middleware (presented in chapter 4) does not claim to solve all the challenges inherent in an ambient ecology, but at least the important challenges and their relationships are to be addressed.
1.8 Research Methodology

The general approach has been to design a conceptual platform (egocentric interaction) based on existing ideas from literature in the field, application area and other sources that promise novel innovation possibilities. Refer to Fig. 1.3. Such conceptual ideas were transformed into system design, followed by prototype development. The prototypes (different components of the easy ADL ecology like the activity recognizer, the situative space tracker, etc.) usually evolved over several rounds of iteration, followed by an evaluation of the easy ADL ecology as a whole. The evaluations have focused on both the technical and the technological aspects of the prototype (the easy ADL ecology) and the conceptual aspects (related to egocentric interaction).

**Fig. 1.3.** The *de facto* design process/research method. The larger arrows denote the most influential development cycle, the smaller arrows denote (re-)design paths less frequented, and the dotted arrows show how external information, tools and knowledge have been incorporated into and used in the design. Adapted from (Pederson, 2003).

The methodology used in exploring a novel interaction paradigm for ambient intelligence is more of a breadth first search than a depth first search for the following reason: to address the challenges in the field of ambient intelligence from a birds-eye view taking the picture as a whole instead of focusing on specific challenges in ambient intelligence ignoring the relationships among those specific challenges. In this context, specific challenges for instance could be activity recognition, object location tracking,
developing ambient intelligence applications, facilitating situated interaction, etc. Such an approach is uncommon and PhD theses usually focus on specific challenges. As mentioned earlier, a field like ambient intelligence is multidisciplinary requiring research efforts on several interrelated research topics in parallel, thereby introducing a need to look at the landscape of ambient intelligence as a whole where the whole is much bigger than sum of its parts.

The evaluation environments used in this work for establishing “proof-of-concept” are of two types: (1) an immersive virtual-reality home environment to ignore the technological challenges and to focus on the conceptual and technical challenges; and (2) a living laboratory home environment to address the technological challenges and to improve on the conceptual and the technical challenges. In-situ environments probably will yield results that are externally valid (Mitchell and Jolley, 2001) in comparison to the virtual reality environment or the living laboratory environment. Since the research work is exploratory in nature with limitations in existing technologies like sensing, networking, interfaces, etc., an in-situ environment would mean a step too far in the future.

1.9 Contributions

This explorative thesis raises several novel issues in developing an interaction paradigm for ambient intelligence. The main area of work is on human-environment interaction, ambient ecology, smart objects, and context-aware computing. The main contributions are:

- **Development of a conceptual platform referred to as Egocentric Interaction** for modeling and facilitating human-environment interaction within an ambient ecology. A body-centered situative space model based on proximity, human perception and action possibilities is described that challenges the traditional device-centric view of facilitating human-computer interaction by replacing the traditional concept of input and output with human perception and action. Also, taking a physical-virtual design perspective (Pederson, 2003) allows for treating both the physical and the virtual artifacts, situations, and activities alike removing the bias towards the physical or the virtual aspects enabling better integration of the two realms.

- **Development of an infrastructural platform referred to as the easy ADL ecology** for the realization of ambient intelligence. Several smart objects with varying capabilities are designed and developed along with a personal activity-centric middleware for handling four important challenges within an ambient ecology namely: (1) tracking and managing smart objects; (2) tracking and monitoring a human agent’s situation; (3) recognizing the human agent’s activities and actions; and (4) facilitating egocentric interaction with ambient intelligence applications. The infrastructural platform eases the development ambient intelligence
applications by making it feasible for the application developers to ignore typical challenges within ambient intelligence.

- **Development of living laboratory and immersive virtual reality simulation smart environments** for conducting further research in ambient intelligence.

### 1.10 Thesis Outline

This thesis is aligned as follows:

- **Chapter 2 (Emerging interaction paradigms and approaches):** An overview of the novel interaction paradigms for ambient intelligence like tangible user interfaces, surface computing, ambient displays, multimodal interaction, perceptual user interfaces, affective user interfaces, activity-based computing, reality-based interaction, and embodied interaction are presented followed by an overview of the existing infrastructural possibilities for ambient intelligence using smart objects, wearable computers and smart environments.

- **Chapter 3 (Egocentric Interaction: a human-centered interaction paradigm):** An overview of the cognitive science theories describing human cognition and the cognitive architectures developed in the field are presented. Egocentric interaction as a theoretical framework for modeling and building ambient ecologies is described in detail. The principle of egocentric interaction like situatedness, the physical-virtual equity principle, the proximity principle, perception and action instead of “input” and “output”, etc. are presented. A physical-virtual design perspective for ambient intelligence and a situative space model for describing a human agent’s situation are described.

- **Chapter 4 (The easy ADL ecology: an infrastructure for ambient intelligence):** The easy ADL ecology comprising of smart objects, a personal activity-centric middleware, ambient intelligence applications and a human agent in the middle of it all is described in detail. The easy ADL ecology is initially developed in immersive virtual reality and later transferred to a living laboratory setup. The infrastructure handles important challenges within ambient ecologies like managing smart objects, tracking a human agent’s situation, recognizing their activities and facilitating interaction with ambient intelligence applications. The interaction management rules and the interaction techniques used in the easy ADL ecology are described as well.

- **Chapter 5 (User experience evaluation of the easy ADL ecology):** The importance of user experience evaluation for ambient ecologies is discussed followed by a detailed description of the evaluation setup and the qualitative results obtained. The user experience evaluation is also a “proof-of-concept” for the theoretical concepts of egocentric interaction based on which the easy ADL ecology is developed.
Chapter 6 (Tracking state changes of physical and virtual objects): This chapter describes the efforts taken in tracking the state changes to physical and virtual objects. A wireless sensor network based on the smart objects in the easy ADL ecology is used for tracking the state changes to physical objects. Virtual objects change their state based on manipulation using speech and gesture modalities. Speech and gesture tracking systems and their accuracies are briefly presented.

Chapter 7 (Situative space tracking across multiple modalities): A situative space tracking system based on wireless LAN signal strength measures is presented in detail followed by other complementary approaches to tracking the situative spaces spanning across multiple modalities like visual, audial, speech and gesture.

Chapter 8 (Recognizing human activities and actions): Two activity recognition systems built based on the situative space model is described in detail. The information channels for activity recognition obtained from situative space tracking have yielded promising results in an immersive virtual reality simulated easy ADL ecology.

Chapter 9 (Discussion and conclusion): Egocentric interaction as an interaction paradigm for ambient intelligence is discussed followed by a summary of the thesis, the contributions, the limitations and the future work.

1.11 Publications

Journal and Magazine

- Surie, D., Pederson, T., Janlert, L-E.: "The easy ADL home: A physical-virtual approach to domestic living". Journal of Ambient Intelligence and Smart Environments, Smart Home Thematic Issue, IOS Press, 2, 3 (August 2010), pp. 287-310. (Surie et al., 2010b)

Conference


**Workshop**


**Book Chapter**


**Keynote Speech**


Chapter 2

Emerging Interaction Paradigms and Approaches

This chapter presents emerging interaction paradigms and approaches in the context of ambient intelligence. In particular, tangible user interface, ambient displays, context-aware computing, reality-based interaction, embodied interaction, and activity-based computing are investigated briefly and compared to the proposed egocentric interaction in terms of: (a) human-environment interaction; (b) human-centered interaction; (c) perception and action instead of input and output; and (d) physical-virtual equity. A survey of the existing infrastructure for the realization of such emerging interaction paradigms by augmenting computing technology with everyday objects (i.e. smart objects), everyday environments (i.e. smart environments) and wearables (i.e. wearable computers) is described in detail.

2.1 Traditional Interaction Paradigms

The term Human-Computer Interaction (HCI) refers to the study of interaction between humans and computers. The scientific field of HCI has seen considerable advancements in the last 30 years, especially since the publication of (Card et al., 1983). Most of the research efforts in HCI attempt to make use of human skills and abilities that are naturally developed by leading a life in this physical world.

2.1.1 Command-Line and Textual Interfaces

Initial interfaces are command-line interfaces where the user of a computer sits in front of a terminal screen, and enters a specific command to perform specific tasks and wait for a reply. Only a part of the terminal screen real estate is used by command-line interfaces. An extension of command-line interfaces is the textual interfaces that attempt to make use of the entire terminal screen real estate. Both command-line interfaces and textual interfaces are developed based on the typewriter metaphor making use of human skills and abilities in language and typing. The notion of *interactive loop* and *dialogue* were introduced.
2.1.2 Graphical User Interfaces

Graphical user interfaces (GUIs) are introduced as an attempt to visualize abstract computational entities as virtual objects that could be directly manipulated (Shneiderman, 1983). The attempt is not to simply replace words by icons, but to introduce a two-dimensional space that mirrors a simplified model of the physical world inhabited by human agents. All that is important for a human agent is assumed to happen within the two-dimensional screen referred to as virtual desktop. The mouse was introduced to facilitate direct manipulation of virtual objects. GUIs lead the way towards the popular WIMP (Windows, Icons, Menus and Pointers) interaction paradigm, an inherent part of desktop computing developed based on an office or desktop metaphor (Smith et al., 1989). Interaction paradigms usually include important design examples and use scenarios, important techniques and technologies, key problems and challenges, articulations of ideals and goals to pursue, interpretations of key concepts, such as ‘user,’ ‘interface,’ ‘interaction,’ and finally, groups or communities of people (researchers and interaction designers) developing and defending the paradigm. Since the WIMP interaction paradigm is based on a simplified model of the physical world and restricts a human agent’s perception and action capabilities to screen, mouse and keyboard, it does not support the wide range of human abilities and skills exhibited by human agents in interactions with the physical world. Also, the WIMP interaction paradigm assumes that a human agent could devote their complete attention in interacting with the desktop computer, and that there is nothing else that is important enough to attract a human agent’s attention in the physical world. This could be associated to the fact that the WIMP interaction paradigm operates in a closed world and ignores a human agent’s physical, social and cultural context. The WIMP interaction paradigm is also often restricted to office type of activities performed in the virtual world (Pederson, 2003), while support for the physical activities performed by human agents is ignored.

With the advancements in computer technology like processing power, memory, storage, communication and with the advent of novel sensors, actuators, interface and interaction technologies, research efforts began in the early 1990s in pushing computers out of its desktop shell. Two types of research efforts began: (1) exploring novel interaction paradigms and approaches beyond the WIMP interaction paradigm; and (2) exploring suitable infrastructures and model of computing for facilitating such novel interaction paradigms beyond desktop computing.

2.2 Emerging Interaction Paradigms

Many research efforts in exploring post-WIMP human-computer interaction have begun by focusing on various human capabilities, skills and limitations commonly observable in the physical world. The interaction paradigms to be presented in this section are a collection of popular and well-established
Emerging Interaction Paradigms and Approaches

paradigms, while there are other interaction paradigms with similar focus or less-established those are ignored in this survey.

2.2.1    Tangible User Interfaces

Tangible User Interfaces (Ishii, 2008) (also referred to as Graspable User Interfaces (Fitzmaurice et al., 1995)) allow a human agent to directly interact with physical artifacts with computing capabilities, instead of indirectly manipulating virtual objects using a pointing device. The human agent is not restricted to two-dimensional virtual object manipulations typical with WIMP interaction paradigm. The virtual object manipulations are extended to a three-dimensional space with the possibilities of using physical artifacts (representing abstract computational entities). Tangible user interfaces are based on human abilities to physically touch, grab, hold and manipulate everyday objects in the physical world with ease. Tangible user interfaces extend the design space in terms of shape, size, color, weight, and texture in comparison to traditional graphical user interfaces.

Typical research in tangible user interfaces focus on building special purpose systems using physical laws and object manipulation capabilities, while smart objects (Kawasar et al., 2008) to be described later focus on making the existing physical objects in the environment smart by augmenting technology. Tangible user interfaces include a wide range of physical objects that allow physical manipulations to interact with computers. However, tangible user interfaces usually focus on explicit input from a human agent while it is possible to obtain implicit input (Schmidt, 2000) using the tangible user interface infrastructure. Tangible user interfaces attempt to integrate the physical and the virtual world similar to the visions of egocentric interaction. However, the physical and the virtual aspects are not (not necessarily) uniformly handled at the artifact, situation and activity levels, while physical-virtual equity is an important principle of egocentric interaction.

Tangible Bits (Ishii and Ullmer, 1997), refer to the notion of giving physical form to digital (or virtual) information. Here physical representations are computationally coupled to virtual objects, and their physical states to the virtual objects’ states. The attempt is to include computation naturally within everyday environments using human experience in dealing with physical phenomena. Tangible user interfaces embody physical representations that are both directly perceivable and manipulable by human agents. While tangible user interfaces usually ignore keeping track of if a physical representation is perceivable or manipulable by a human agent at a particular moment in time, egocentric interaction focuses on both physical and virtual representations that are perceivable and manipulable by a human agent at a particular moment in time. Refer to the situative space model in chapter 3.

Tangible user interfaces incorporate additional novel user groups, allowing for non-professionals, children, and elderly to interact with computers similar to the visions of egocentric interaction. While work on
tangible user interfaces has produced many promising special-purpose systems based on tailor-made real-world tokens (Ullmer et al., 2005) connected to virtual representations, egocentric interaction includes “ordinary” physical objects already inhabiting everyday environments. Another difference is that while egocentric interaction embraces mobile computing, most work on tangible user interfaces to this date does not.

2.2.1.1 Tangible User Interface Prototypes

Refer to Fig. 2.1 for the classical examples of existing tangible user interfaces. The marble answering machine invented by Durrell Bishop associates incoming voice messages to marbles. By manipulating the marbles, a message can be played back, can inform the caller to call back, delete messages and also organize them. Media blocks allow for physically capturing, retrieving and manipulating digital media like images, video, etc. Bricks are graspable objects for providing 6D inputs and its GraspDraw application allows a human agent to select or delete rectangle, triangle, circle, etc. with different colors using a physical tray and an ink-well metaphor.
Music blocks from Neurosmith are a learning toy for kids that enable them to create musical compositions by playing with the blocks.

Lumino (Baudisch et al., 2010) is a more recent prototype made of tangible blocks with a large number of thin glass fiber bundles, specifically useful for tabletop computers. The glass fiber blocks are transparent and optically reflect light onto the table surface with its marker. Such optical reflections are captured by a camera. The blocks are battery-free and they allow light reflection through several blocks stacked on a table surface. Lumino can be used as a three-dimensional construction kit where the blocks could be stacked and arranged on a table. Lumino evaluates the simulated three-dimensional architectural structure in terms of its stability, and provides assistive information and reports unstable structures. Microsoft’s surface computer is used for the prototype development. Refer to fig. 2.2.

![Figure 2.2](image)

**Fig. 2.2.** The lumino construction kit application: (a) recognizes and labels individual components; (b) when two components are merged with a bridge the construction kit labels them as one; (c) lumino warns about an unstable structure; (d) the markers used are shown (Baudisch et al., 2010).

### 2.2.2 Surface Computing

Surface computing refers to interaction techniques that take place on or in-front of an interactive surface. Tabletop displays and wall displays are
common forms of surface computing. Surface computing prototypes are similar to specific set of tangible computing prototypes that makes use of a surface computer like the Lumino (Baudisch et al., 2010). Tangible user interfaces typically focus on physical objects and their manipulation while surface computing focuses on interaction with an interactive surface. The different interaction techniques supported by surface computing include:

1. **Multi-touch interaction on an interactive surface**: single and multiple human agents make use of their hand fingers (both hands) in experiencing multi-touch interaction. SmartSkin (Rekimoto, 2002) supports multiple hands and multiple human agents interaction with an interactive table.

2. **Tangible objects on an interactive surface**: this is similar to tangible user interfaces where the tangible objects could be manipulated on the surface for interacting with computers. An example is Lumino (Baudisch et al., 2010).

3. **Tangible objects near an interactive surface**: here the tangible objects are manipulated in front of (or) above an interactive surface for interacting with computers. Since the interaction take place in a space near to the interactive surface, the tangible objects could be manipulated in 3D free space. An example using 3D position manipulation of a pen above an interactive surface is described in (Subramanian et al., 2006).

4. **Hand gestures near an interactive surface**: this technique allows human agents to provide gestures, typically hand gestures even though head movements can also be used in 3D free space near an interactive surface for interacting with computers. The concept of continuous interaction space described in (Jota et al., 2011) is an example for this interaction technique.

### 2.2.2.1 Surface Computing Prototypes

SmartSkin (Rekimoto, 2002) is an early surface computing prototype that supports capacitive sensing with a mesh-shaped antenna for determining the hand positions and the distance between the hand positions to the surface. The mesh density can be increased to track hand positions with finger-level granularity. This approach does not suffer from occlusion and lighting problems and enhances the possibility to manipulate virtual objects.

An attempt to combine the on-surface and in-front of surface interaction techniques to provide a continuous interaction space is described in (Jota et al., 2011). Such an approach allows for interaction with virtual objects while touching the surface and continues with the virtual object manipulation when the hands are no longer touching the surface and gestures are provided in the free space. Refer to Fig. 2.3. The continuous interaction space approach has several applications involving browsing namely the digital book for touch-
based and gesture-based flipping of book pages, and the photo viewing application for browsing through a stack of photos as shown in Fig. 2.6. Infrared emitters on a human agent’s gloves in combination with infrared cameras are used to track hand gestures and object manipulation in front of the surface while a touch sensitive surface is used for tracking the human agent’s touch-related gestures.

![Continuous Interaction Space and Touch Surface](image)

**Fig. 2.3.** A continuous interaction space above the interactive tabletop surface (left) and browsing photos by moving a tablet computer above an interactive surface (Jota et al., 2011).

The continuous interaction space is similar to the concept of perception space and action space to be described in chapter 3. Everything of interest happens to take place within the perception and the action spaces similar to the continuous interaction space. The difference is that the continuous interaction space is centered on the interactive surface while the action space and the perception space are centered on a human agent’s body and their situation. Within the continuous interaction space, the length of the virtual stack is determined by the human agent’s arm reach similar to describing the action space boundary along the touch modality. Also the tablet computer’s orientation limits the photo displayed on it which is similar to describing the perception space boundary along the visual modality. While human perception and action capabilities are the starting point for egocentric interaction, the continuous interaction space is device-centric. The continuous interaction space has a tracking range of 1 meter above the surface while the tracking range of action space along the touch modality (at least in theory) is determined by a human agent’s physical arm reach while being stationary. The concept proposed in continuous interaction space is not restricted to horizontal tabletop displays, and could well be generalized to vertical surfaces like an interactive wall or an interactive white board.

Another approach where the space above the surface is divided into different interaction layers with the possibility to use the 3D position of a pen in specific layers to facilitate interaction is researched in (Subramanian et al., 2006). Here, the physical space is divided into different interaction layers, while a similar approach of dividing the situative spaces (refer to chapter 3) into perception space, action space, recognizable set, examinable set, etc. but
based on differences in human perception and action capabilities is addressed in egocentric interaction. The device-centric approach in (Subramanian et al., 2006) forces a human agent interacting with virtual content to place the pen in specific heights above the surface which is not natural and often result in unintended drifts from one layer to a different layer. By taking a human and more specifically their body-centered approach, setting up the boundaries for perception space and action space more natural and according to that human agent’s capabilities.

There are several commercial surface computers that are available including Microsoft Surface (Wall, 2009) and DiamondTouch (Dietz and Leigh, 2001). Microsoft surface supports direct interaction with its surface through multi-touch and natural finger manipulations. The physical objects on the surface are tagged and are recognized by the surface computer allowing tangible interaction. SurfaceWare (Dietz and Eidelson, 2009) uses Microsoft surface with dynamic tags that are passive and detect not only the tagged objects but also their state changes. By supporting interaction with multiple human agents simultaneously, Microsoft surface supports computer supported cooperative work (CSCW).

2.2.3 Ambient Displays and Informative Art

Ambient displays focus on turning the space around human agents into an ambient interface where human agents could perceive virtual objects not only through their central attention, but also through their peripheral attention (Wisneski et al., 1998). This idea of ambient media where a human agent’s peripheral awareness is used to complement their central attention in designing interfaces common within ambient displays and tangible user interfaces is borrowed as an inspiration within egocentric interaction.

The scope of ambient displays is to convert the entire physical environment into an interface to virtual objects. Ambient spaces could present information through subtle changes like variations in lighting condition or in the sound levels in an environment. Nature is expressive and presents information in multiple modalities of which some are experienced by human agents without explicitly concentrating on such information. For instance, a couple taking a walk in the nature might experience the wind, the sound of birds, and fresh air, and might still be pre-occupied with their conversation which is their primary activity of focus.

Environments occupied by human agents offer implicit cues that are perceived through their peripheral attention. In this way, the central attention is not overloaded and more information can potentially be presented in the periphery of human attention. WIMP interaction paradigm ignores the concept of ambient media and focuses on presenting information usually with high density on a small rectangular screen or windows. The various applications attempt to get its share of the limited real estate on the screen thereby overloading a human agent’s attention. Ambient displays on the contrary are intended to avoid overloading of a human agent attention. Such
Emerging Interaction Paradigms and Approaches displays can find their place in home, office, shopping mall, business center, etc., and are useful in moving towards the vision of ambient intelligence.

The multiple modalities in which ambient displays present information are often device or the specific ambient display-centered. While it is not a problem in itself, it greatly restricts the multiples modalities through which information could be effectively presented to a human agent based on their capabilities in a particular situation. Also, typical ambient displays do not keep track of a human agent’s attention in being adaptive to their ambient presentation.

Ambient displays are part of the concept of tangible bits (Ishii and Ullmer, 1997) where the aim is to integrate the physical and the virtual worlds by turning the ambient space around human agents into an interactive space for both perceiving virtual information and manipulating virtual information. While tangible user interfaces usually focus on facilitating virtual information manipulation similar to the concept of action space and ambient displays focus on facilitating virtual information presentation similar to the concept perception space, egocentric interaction addresses perception and manipulation of both physical and virtual objects. In that sense, the concept of tangible bits are restricted to virtual information and physical representation of virtual information while egocentric interaction includes physical objects, virtual representation of physical objects, virtual objects and physical representation of virtual objects.

2.2.3.1 Ambient Display Prototypes

To begin with the classical prototypes, the water lamp (Wisneski et al., 1998) or an ambient fixture changes the light patterns projected onto the ceiling of a room depending on changes in the virtual world. The water lamp contains a pan of water with solenoids mounted on it. By actuating the solenoids that tap on the surface of the water, water ripples are created. The light that is refracted through the pan of water creates patterns on the ceiling. The pinwheels (Wisneski et al., 1998) are also ambient fixtures that use airflow metaphor in presenting virtual information. Controlling the airflow in a room is reported to be challenging and has led to varying the speed of pinwheels for presenting information. The pinwheels are made up of folded fiberglass and are controlled by a motor. Ambient fixtures can present information about geographical events, atmospheric conditions, etc. in an aesthetically pleasing manner.

The CareNet photo frame display (Consolvo et al., 2004) is an ambient photo frame display useful for elderly people in their home environment to get information about their everyday life and to keep in touch with caregivers. There are two modes: (a) in the ambient mode, CareNet photo frame display contains seven icon types namely meals, medications, outings, activities, mood, falls, and calendar with the possibilities of several icons of the same type, for instance three meals icons could represent breakfast, lunch and dinner; and (b) in the interactive mode, the icons could be manipulated by touch to get further virtual information or services. The CareNet photo frame
display obtains information through sensors, the elderly person and the caregivers. Virtual services include a calendar that is self-editable by the elderly, and includes the person’s appointments and the available transportation to meet the appointments. The CareNet photo frame display was not perceived as a computer but as an everyday object with additional virtual functionalities and aesthetically pleasing design, blending well within the different everyday home settings.

Fig. 2.4. Examples of Ambient Displays and Informative Art.

The Gossip wall (Streitz et al., 2003), part of the Ambient Agoras project focuses on situated interaction within office environments facilitating informational needs of individuals and groups. The proximity of human agents with reference to the gossip wall determines the “zones of interaction” in which the individual human agents are currently situated. The three zones of interaction are ambient zone, notification zone and interactive zone. Human agents who are situated far away from the display and are not tracked by the sensors embedded on the wall are said to be in the ambient zone where general information is presented. Human agents who are close to the wall are in the notification zone where personal information to an individual or a group is presented. In the interaction zone, human agents are very close to the wall and can personally interact with the wall and
manipulate virtual information. The gossip wall is equipped with 2 RFID readers with different ranges and a wireless LAN network. The display contains 124 cells with each cell comprising of a cluster of LEDs with the possibility to adjust their light intensity.

The concept of using proximity in defining the interaction zones is similar to egocentric interaction in two ways: (a) egocentric interaction is based on the proximity principle (refer to chapter 3) where objects that are of close proximity falls into the action space “interaction zone” while objects that are far away falls in the world space “interaction zone” with other spaces like perception space, recognizable set, etc. that can be represented in terms of proximity (refer to the WLAN signal strength-based situative space tracking system in chapter 7); and (b) egocentric interaction facilitates situated interaction using the situative space model (refer to chapter 4). While the Gossip wall is device-centered where proximity is measured with reference to the gossip wall, egocentric interaction is body-centered where proximity is measured with reference to a human agent’s body.

Informative art (Redström et al., 2000) is computer augmented or amplified art work that dynamically reflect virtual information in an aesthetical and interactive manner using traditional art objects like paintings and posters. Informative art artifacts could be considered as a special branch of ambient displays that combines the idea of using artwork to represent virtual information. Similar to ambient displays, informative art artifacts are expected to be an integrated part of the destined environment. Informative art usually represent abstract information inspired by some existing piece of art as shown in Fig. 2.5. The arrival and departure times of two bus lines are represented using a composition of rectangular color fields (red, blue and yellow) together with black lines, also referred to as Mondrian art named after the Dutch artist Piet Mondrian.

![Fig. 2.5](image_url) Arrival and departure times of two bus lines represented like a Mondrian painting (Redström et al., 2000).
2.2.4 Multimodal Interaction

Human agents communicate with other human agents in an environment using several modalities, often in parallel and in a natural manner. The current situation of a human agent greatly determines the modalities that could be more effective in comparison to others. Multimodal interaction has a long history going back to the era of desktop computing and is useful in moving towards the visions of ambient intelligence.

Modalities like speech, hand gestures, facial expressions, vocal emotions, speech output and tactile feedbacks are more natural to a human agent in comparison to modalities like visual display in the form of a screen, mouse and keyboard. Taking inspiration from everyday environments and human-human interaction in such environments, research efforts have focused on using gestures, hand posture and gaze (Kisacanin et al., 2005); facial expressions (Fasel and Luettin, 2003), vocal emotions (Oudeyer, 2003), speech, and combining multiple modalities (Jaimes and Sebe, 2007) in facilitating human interaction with digital information. Human emotions are a particular modality actively researched in the field of affective user interfaces (Hudlicka, 2003).

Multimodal interfaces are expected to be context-aware (Dey, 2001) in automatically selecting the input and output modalities that are ideal for different situations. Multimodal interfaces could be used independently, but an important feature of multimodal interfaces is that they combine the different modalities such that the modalities are coordinated both spatially and temporally. Such co-ordinations in practical application of multimodal interfaces are a challenge to address.

Egocentric interaction supports interaction distributed across several modalities. The situative space model (refer to chapter 3 and chapter 7) is centered on a human agent’s body and spans across different modalities through which the human agent can perceive and act in an environment. While the typical multimodal interfaces focus on specific modalities, often driven by individual devices in hand which might compete for human attention through similar modalities. Egocentric interaction starts with the capabilities of a human agent across different modalities as a starting point and then checks for the devices that supports those modalities. Taking a human-centered stance enables facilitating human-environment interaction comprising of multiple devices distributed across multiple modalities in an orchestrated manner.

2.2.5 Affective User Interfaces

Affective User Interfaces (Picard, 1997) focus on using a human agent’s emotional cues as important context information for interacting with computing systems. Affective user interfaces keep track of the human agent’s affective state and adapt their behavior to provide an emotional response that is appropriate and respectful to the human agent. Egocentric interaction does
not include human emotions at present, but it may be an interesting theme to explore in the future.

Human agents are good in recognizing another human agent’s affective state based on their speech tone, facial expressions, body posture and motion, etc. and adapt their social interaction with them. Interfaces for ambient intelligence should respect a human agent’s affective state for providing positive user experiences within such environments.

In (Gunes and Piccardi, 2009), automation recognition of a human agent’s affective states is performed by using the visual cues from the human agent’s facial expressions, and the cues obtained from the human agent’s hands and shoulder movements. This work focuses on fusing multiple modalities considering the modality dynamics in recognizing emotions like happiness, anger, uncertainty and puzzlement.

In (Gilroy et al., 2011), physiological signals are used as a reference to measure a human agent’s emotional states. Surface electromyography (EMG) and galvanic skin response (GSR) signals known to correlated to specific dimensions of emotion (pleasure and arousal). They are compared to real time continuous values of these dimensions obtained from affective multimodal fusion.

2.2.6 Perceptual User Interfaces

There are several approaches to interaction that focus on perceiving and modeling a human agent for the purpose of developing a mode of interaction that is more human-centered, adaptive and implicit. Perceptual user interfaces (Turk and Robertson, 2000) are developed based on how human agents interact naturally with other human agents and with the physical world. It is characterized by: (a) understanding natural human capabilities like communication, motor, cognitive and perceptual skills; (b) approaches to capture and make use of such capabilities using machine perception and reasoning; and (c) incorporate multiple modalities for natural human-computer interaction.

Perceptual user interfaces are closely related to egocentric interaction in the sense that egocentric interaction is also based on understanding human capabilities like perception, action, attention and intention. While perceptual user interfaces are more general as a concept, egocentric has specific models like the situative space model which is centered on human body. Possessing specific models takes a step closer to implementation and is more useful for designers. While perceptual user interfaces consider human expectations and the semantic nature of human interactions, egocentric interaction for the moment has not dwelled into such issues. Another difference is that perceptual user interfaces focus on facilitating human-like perceptual capabilities within computers in knowing a human agent’s facial expression, hand operations, body posture, etc. in facilitating interaction. However, egocentric interaction focuses on capturing what a human agent possibly perceives and potentially acts in a situation for facilitating interaction.
Perceptual user interfaces track information about a human agent from the “second” perspective while egocentric interaction track them from the so-assumed “first” or “egocentric” perspective. Perceptual user interfaces typically focus on using machine understanding of human capabilities for facilitating interaction with virtual objects and for supporting human virtual activities while egocentric interaction supports interaction with physical and virtual objects and supports physical, virtual and physical-virtual activities (Pederson, 2003).

2.2.7 Attentive User Interfaces

Attentive user interfaces (Vertegaal, 2003) refer to computing interfaces that are sensitive to a human agent’s attention. They attempt to measure and model human attention and use this information to structure their communication with the human agent. (Vertegaal et al., 2006) propose five key properties of attentive systems: (1) to sense attention, (2) to reason about attention, (3) to regulate interactions, (4) to communicate attention, and (5) to augment attention. Such properties could well be applied to systems built based on egocentric interaction with the specific focus on human attention be shifted to a particular human agent as a whole.

The WIMP interaction paradigm assumes that a human agent spends all their attention on interaction with the desktop computer. In an ambient ecology, a human agent would be surrounded by multiple computing devices that are available 24x7 willing to interact with a human agent and eventually placing strain on their attention capabilities. Since human attention is a limited resource, it should be handled efficiently. Human agents in a physical environment attend to different objects and events with varying levels of attention optimal to the intended situation and their attention capabilities.

Attentive user interfaces typically focus on tracking a human agent’s attention, while egocentric interaction does not keep track of a human agent’s attention explicitly. However, some characteristics of human attention like user presence, proximity, orientation, speech, gaze, etc. are implicitly included in the situative space model, and by keeping track of their activities, it is possible to associate human attention to the situative spaces as described in chapter 3. Attentive user interfaces are usually device-centric addressing human attention in isolation in comparison to handling it as a whole centered on a human agent as in egocentric interaction. Attentive user interfaces mediate the amount of attention required for interacting with virtual objects by presenting virtual objects that move between the central attention and the peripheral attention of a human agent similar to ambient displays and egocentric interaction. Attentive user interfaces could be considered as a specific case of perceptual user interfaces and context-aware interfaces where the primary focus is on human attention (Roda and Thomas, 2006, Vertegaal et al., 2006).
2.2.8 Context-Aware Computing

Human agents interact within physical environments taking into account the contextual information, i.e. information that characterizes the human agents’ situations. They communicate rich amount of information implicitly in their interaction with other human agents. The WIMP interaction paradigm ignores a human agent’s contextual information, assuming their situation to be always constant. However, such an assumption is not suitable for ambient intelligence. Context-aware computing (Schmidt, 2002) attempt to sense, recognize and infer contextual information, and use such information to enhance the human agent’s interaction with computing systems. Context includes a human agent’s location, activity, proximity to objects, etc.

Contextual information could be considered as implicit input, i.e. input that is not directly intended to be provided to a computing system, thereby increasing the bandwidth of communication between humans and computers. Context-aware computing focuses on using contextual cues available in the physical world as a means of improving the richness of communication between human agents and computing systems. According to (Dey, 2001), important features of context-aware application include: (a) presentation of information and services; (b) automatic execution of services; and (c) for tagging context to available information.

The term “context” has several interpretations. According to (Schilit and Theimer, 1994), context refers to location, identities of nearby persons and objects, and changes to those objects. Egocentric interaction keeps track of the identities of nearby objects and their state changes in reference to a human agent’s body using the situative space model, apart from including additional contextual information like the human agent’s activity, their capabilities in perceiving and acting upon objects, attention-relationship to the situation, etc. making egocentric interaction context-aware by default. According to (Dey, 2001), “context is any information that can be used to characterize the situation of an entity (i.e., a person, a place or an object) that is considered relevant to the interaction between a user and an application, including the user and applications themselves”. And, “a system is context-aware if it uses context to provide relevant information, and/or services to the user, where relevancy depends on the user’s task”. The importance of including human activity as contextual information is highlighted in this definition.

Typical context-aware systems treat the physical world as subordinate to the virtual world: a human agent interacts with a device and what happens in the surrounding physical world is just considered as contextual cues, not as important as what is going on in the virtual world provided by the device and the way such information affect the physical world. In contrast, egocentric interaction takes an egalitarian stance with regard to the physical and the virtual aspects of a human agent’s situations. Modeling and learning situations for context-aware services in a smart home environment is described in (Brdiczka et al., 2009a) where a situation model comprising of the environment, the human agents within it and their activities is considered and a framework for acquiring and evolving multilayered situations is proposed.
From a context-aware computing (Schmidt, 2002) perspective, the work presented in this thesis typically focuses on user context, while the environmental context and the activity context are presented as well, but with reference to the user context. Refer to the 3-D context model described in (Schmidt, 2002). For more information on context-aware systems, refer to (Baldauf et al., 2007).

2.2.9 Implicit Human-Computer Interaction

Implicit human-computer interaction (iHCI) (Schmidt, 2000) refers to the actions performed by human agents that are not primarily intended to interact with a computing system but is interpreted by an intelligence computing system as implicit input. The term iHCI is more to represent the fact that human agents are not explicitly interacting with computing systems. There has been some criticism of this term, especially since it has been difficult to define what is implicit output if we would still like to stick with the terms input and output from a HCI perspective. iHCI is based on two main concepts namely perception and interpretation. Usually iHCI complements explicit interaction, especially within an ambient ecology. Context-aware computing is an important tool in facilitating implicit interaction (Schmidt, 2000). In an environment where many applications are trying to interact with a human agent, interfaces that are purely explicit will cognitively overload the human agent. Egocentric interaction uses implicit interaction, but it is not the main focus of research.

2.2.10 Activity-Based Computing

Human agents’ interaction within a physical environment is largely driven by their activities. The WIMP interaction paradigm is designed based on an application-centered or document-centered model suitable for specific tasks, but does not provide support for higher-level activities. However, activity-based computing is useful for ambient intelligence.

Activity-based computing (Bardram, 2007) consider human activities as a starting point in facilitating human interaction with computing systems. It attempts to provide computational support at the activity level by aggregating resources required for individual activities. It supports parallel and interleaved activities, and their activity support is context-aware. According to (Li and Landay, 2008), activity-based computing is an approach to interaction that aims at integrating ubiquitous computing technologies with natural human activities.

Activity-based computing uses activity theory (Kuutti, 1996) as a conceptual tool for describing, structuring and analyzing human activities. According to activity-based computing, human interaction with computing systems should be based on the activities performed by a human agent in the physical world, encouraging a mix of implicit interaction (Schmidt, 2000) with explicit interaction. The basic “computational unit” of activity-based
computing is neither files nor applications, but the *activities* of human agents. (Bardram, 2007) lists the following essential principles of ABC:

1. **Activity centered** by selecting a range of services needed to support a human agent’s current activity.
2. **Activity suspension and resumption** to allow a human agent to alternate between several activities.
3. **Activity roaming** to allow for a human agent to be mobile and still obtains activity support.
4. **Activity support** while being adaptive to the resources available.
5. **Activity sharing** for supporting collaborative human activities.
6. **Context-awareness** for recognizing activities in varied context.

Activity-based computing has an ambition of weaving ubiquitous computing technologies into everyday human activities fits egocentric interaction, as well as the support of coexistent implicit and explicit interaction. Egocentric interaction has its primary focus on individual human activities, not on activity sharing, but is otherwise informed by activity theory, as well as by situated action theory (Suchman, 1987), which assigns great importance to the human agent’s situation for shaping the agent’s activities, and by distributed cognition theory, which recognizes that human activities are not confined to the body and brain of a human agent, but involves the environment, the artifacts used and the exchange of information between them. (Bardram, 2007) proposed a simplified view of human activities to be a collection of computational services. Egocentric interaction acknowledges this simplified view, but does not restrict human activities both in modeling and for providing support to just a collection of computational services.

### 2.2.11 Embodied Interaction

Human agents are embodied beings in nature, i.e. they have a physical body and are manifested in the everyday world. Embodiment is beyond physical embodiment. Human agents perform activities within a setting where the setting is not a mere background, but a fundamental and constituent component of the activities providing meaning and value to their activities. Embodied interaction (Dourish, 2001) is based on the fundamental principle of embodiment, and considers the physical and social aspects of interaction. Embodied Interaction (Dourish, 2001) is based on theories of embodied cognition (e.g., (Lakoff and Johnson, 1999)) and the concept of embodiment, emphasizing the interweaving of the physical and the social aspects of interaction. Embodiment is the property of being manifest in and of the physical environment such that the realm of ideas is replaced by the realm of everyday experience. Embodiment is not restricted to physical embodiment, but includes the embodied set of relationships (with other embodied agents), actions, assessments and understandings. According to Dourish, the term “embodied interaction” is used to denote that interaction as such is an embodied phenomenon that takes place in a physical world with social set-up
that lends form, substance and meaning to the interaction. While egocentric interaction does not focus on social aspects in its current formulation, neither does it discourage them. For reasons of simplicity, we are currently working on a “one human agent in an environment” approach rather than directly considering “multiple human agents in an environment.”

2.2.12 Dimension Space Model

Attempts to formalize relationships between physical objects have been done within tangible user interfaces research area such as Token and Constraints (Ullmer et al., 2005) and the Cognitive Dimensions framework (Edge and Blackwell, 2006). Egocentric interaction complements these modeling efforts by extending the size of the modeled space to the rim of human perceptual capabilities (not limited to what fits on a table, for instance), by making the modeled space move with the human agent; and by including virtual objects in the space. The dimension space model (Graham et al., 2000) intends to help in measuring how objects compete for attention from a specific human agent performing a specific task at a specific time. Just like egocentric interaction, the dimension space model includes both physical and virtual objects.Unlike the dimension space model, which focuses on perception alone, the proposed egocentric interaction incorporates action as well.

2.2.13 Reality-Based Interaction

Reality-based interaction (Jacob et al., 2008) attempts to provide a conceptual framework for the emerging trends in HCI. In particular, reality-based interaction attempts to focus on the following four themes relating to human agents: 1) Human agents possess common sense knowledge about the physical world in which they live; 2) Human agents are aware of their body and possess skills to control and coordinate it in the physical world; 3) Human agents are aware of their surroundings and possess skills for negotiating, manipulating and navigating within their environment; and 4) Human agents are aware of other human agents in their environment and possess social skills for interacting with them.

The goal is to push HCI towards interaction that more closely resembles or mimics human interaction with “ordinary” objects and the “ordinary” physical world. Reality-based interaction fits egocentric interaction with regard to the focus on the human agent’s body, environment, and the agent’s commonsense knowledge of the physical world, including notions like the persistence of objects, their various states, etc. However, while RBI is quite general and perhaps somewhat vague on details and specifics in its ecumenical ambitions, egocentric interaction aspires to be more specific, thereby promising to be more useful in practice for designers. Egocentric interaction is based on human perception, action, intention and attention possibilities, with a view to be useful in supporting human agents with their everyday activities.
2.2.14 Mixed-Initiative Interaction

Moving from desktop computing to ambient intelligence, computer systems no longer can assume exclusive access to the scarce resource of human attention: they will have to compete with each other and with real-world events. Gathering information about the state of the world using sensors in an implicit manner can increase the number of decisions that can be made without involving the human agent. If explicit human decision-making still is unavoidable, timing and choice of interaction modality are crucial for minimizing the cost. In short, the classic HCI dialog where human agents have the initiative and the system mainly reacts, needs to change towards a more balanced (and complex) mixed-initiative interaction (Horvitz, 1999) dialog to better accommodate for ambient intelligence. Mixed-initiative interaction is supported by egocentric interaction and the interaction management rules described in chapter 4 is based on this approach.
2.2.15 Human-Centered Computing

Human-Centered Computing (Sebe, 2010) focuses on bridging the gap between human agents and computing systems by having human agents as the primary central focus. The focus is on personal, social and cultural context of human agents, often ignored in technology-centered research. Sciences related to human agents like cognitive science and social science are taken as a starting point which is similar to egocentric interaction where the focus is on human agents. Human-centered computing aims for changing the landscape of computing by introducing new methodologies to design and build computing systems, while egocentric interaction focuses on modeling and facilitating human interactions within an ambient ecology.

According to (Jaimes et al., 2007), human-centered computing deals with the understanding of human beings in designing computing systems (similar to egocentric interaction) supports mixed-initiative interaction. Egocentric interaction supports the possible presence of applications that do much of their work without (conscious) input from the human agent, but will occasionally need to ask for additional information in situations that are not necessarily closely related to the information needed in the first place. For this to work smoothly, we need to approach the level of human sensitivity in human–human communication: know when and how to listen, when and how to comment or raise a question, all based on silently measurable cues such as for instance the pose, body language, and actions performed by the dialog partner.

2.2.16 Discussion

Egocentric interaction, to be described in chapter 3, attempts to unify existing research findings relating to a human agent into a single theoretical platform with the possibility to evaluate and refine it. Egocentric interaction deals with the design space of an environment in which human agents live instead of specific interfaces or interaction approaches through which they interact with computers as described in (Winograd, 1997). This is in contrast to efforts like in tangible user interfaces, perceptual user interfaces, attentive user interfaces, speech user interfaces, gesture user interfaces, etc. that attempts to focus on user interfaces and specific micro-aspects of a human agent. It is important to exchange research findings between the two contrasting styles of exploring suitable interaction paradigms for AmI. The former approach gives the big picture, while the latter approach provides depth in addressing the specific parts of the big picture. Egocentric interaction might lack the low-level perfection of some of the existing work, but views the challenge of human-environment interaction as a whole at a high-level. Egocentric interaction addresses issues and challenges at the interaction level, and needs to be complemented by research efforts at the interface level. Reality-based interaction and embodied interaction have a broader scope in comparison to egocentric interaction, but egocentric interaction is sharper and shares
specific and detailed models like the situative space model to be described in chapter 3.

In comparison to device-centered approaches, egocentric interaction takes a human-centered approach inspired from cognitive science theories. Egocentric interaction uses a human agent’s situation, their body and cognitive capabilities and skills in facilitating human-environment interaction. Egocentric interaction is not technologically driven, although technological advancements are important to use it for facilitating HCI. By taking a human-centered perspective on HCI, egocentric interaction is expected to withstand the series of technological advancements over time. Egocentric interaction is based on a human agent’s perception and action capabilities, considered to be a starting point for a human agent’s behavior in the physical world. In this sense, the egocentric interaction is a by-product of basic research on human agents and is valid along the entire reality-virtuality continuum (refer to section 2.7). Even though different infrastructures for post-desktop model of computing introduce different HCI-related challenges, there are some HCI challenges that are more fundamental, centered on a human agent and probably are applicable to different computing infrastructures. Egocentric interaction attempts to address such fundamental HCI challenges allowing room for complementary and more infrastructure-specific research efforts to exist in parallel.

The survey of emerging interaction paradigms for ambient intelligence reflected on several problems that were not addressed in a unified manner.

1. Lack of support for human-environment interaction: Most of the existing work focuses on human interaction with a single device and is typically device-centric.
2. Lack of support for human-centered interaction: Most of the existing work focuses on specific human-related context ignoring their body, situation and current activities as a whole in facilitating interaction.
3. Support restricted to limited input and output modalities: Most of the existing work ignores the possibilities of viewing a human agent’s perceptual and action capabilities as a whole that spans across multiple modalities at varying levels of granularity for presenting and obtaining information.
4. Lack of support for uniformly handling both physical and virtual aspects in facilitating interaction: Most of the existing work does not treat artifacts, situations and activities that spans across the physical and the virtual realm alike.

2.3 Reality-Virtuality Continuum

The emerging interaction paradigms allow a human agent to be immersed in computer-generated virtual environments at different levels. According to Milgram’s concept of reality-virtuality (RV) continuum (Milgram and Kishino, 1994), there exists a continuous scale between completely physical (or real) environments populated with purely physical objects to completely virtual environments populated with purely virtual objects. Refer to Fig. 2.7. The
environments between these two extremes are populated with both physical and virtual objects with varying levels of augmentation. Here the term augmentation refers to the state of being augmented with computing technologies. Mixed-reality environments are in the middle of the RV continuum containing both physical and virtual objects typically with equal augmentation. However, augmented reality takes a bias towards real-world elements while augmented virtuality takes a bias towards virtual world elements. Physical objects usually has the advantage of providing rich affordances (Norman, 1998a) and can be manipulated using natural human skills, while virtual objects has the advantage of being light-weight could be instantly transported across large distances with ease. By allowing for illusive co-existing of physical and virtual objects, mixed-reality environments attempt to combine the best of both the physical and the virtual world. In order to create the illusion of coexistence, virtual objects should be accurately positioned in the physical environment in real time and well aligned with physical objects, which is a major challenge within mixed-reality research.

![Fig. 2.7. Milgram’s Reality-Virtuality Continuum (Milgram and Kishino, 1994).](image)

### 2.3.1 RV Continuum and Ambient Intelligence

At one end of the RV continuum, a human agent is fully immersed into a computer-generated artificial environment researched in the field of immersive virtual reality (Burdea and Coiffet, 2003); and at the other end, a human agent is fully immersed within a physical environment with no computing augmentation. By augmenting many small, invisible computers on physical objects in the physical environment, a human agent in such a physical environment could benefit from computing services typically researched in the field of ubiquitous computing (Weiser, 1995). Egocentric interaction and the work presented in this thesis falls in the mixed-reality zone where the attempt is to allow for the coexistence of physical and virtual objects in moving towards the visions of ambient intelligence. Ambient intelligence, at least in theory is neither biased towards physical environments nor towards virtual environments, and it allows for varying levels of physical and virtual object augmentations. However, typical research
within ambient intelligence address challenges with a bias towards the physical environment.

### 2.3.2 Augmented Reality and Mixed-Reality

Research in augmented reality has the central theme of augmenting everyday physical objects in the physical world with virtual information such that the human agents can take advantage of their skills in interacting with everyday physical objects while benefiting from the support of computing systems. (Mackay et al., 1998) describes three basic strategies used in augmented reality research: augment the human agent, augment the physical object, and augment the environment surrounding the human agent and the object. In section 2.3, the three strategies are described as part of ongoing research on exploring post-desktop computing infrastructures for ambient intelligence. Egocentric interaction to be described in chapter 3 is consonant with these strategies and supports the integration of the physical and the virtual world, but disallows the treatment of the virtual world as subordinate to the real world (physical world), which is characteristic of augmented reality. Egocentric interaction attempts to give equal priority to physical and virtual objects, physical and virtual situations, and physical and virtual activities similar to research in mixed-reality. This approach is also expected to relieve the human agents from additional efforts (jumping between the physical world and the virtual world) needed when performing activities that make heavy use of both the physical and the virtual worlds (Pederson, 2003). Also, egocentric interaction encompasses interaction in every modality accessible to human agents, not just the visual dimension typically targeted in augmented reality systems.

#### 2.3.2.1 Augmented Reality and Mixed-Reality Prototypes

A few related augmented reality and mixed-reality interfaces are briefly presented. Augmented reality guidance for needle biopsies (described in (Azuma, 1997)) where physicians can directly see into a patient as an alternative to ultrasound guided procedures, while car designers are aided in developing proof of concept in the Fata Morgana project (Klinker et al., 2002).

The magic book (Billinghurst et al., 2001) is a mixed-reality interface that overlay three-dimensional virtual models on physical book pages. It attempts to provide smooth transition between reality and virtuality as illustrated in Fig. 2.9. A human agent willing to experience the physical reality by reading a physical book is shown in Fig. 2.9 (left), while the human agent experiencing an augmented reality environment where a physical book is augmented with 3D virtual characters is shown in Fig. 2.9 (center). The human agent within an immersive virtual reality environment can move around freely in the scene as illustrated in Fig. 2.9 (right). Magic book could be used as an interactive story book for children, and for the visualization of geological data.
Fig. 2.8. Augmented reality used in guiding for needle biopsies (left) (Azuma, 1997) and car designing (right) (Klinker et al., 2002).

Fig. 2.9. The MagicBook is used to move between reality and virtual reality. Physical reality is shown in the left, augmented reality in the center and virtual reality in the right (Billinghurst et al., 2001).

Mixed-reality is used for building a rowing simulator (Costanza et al., 2009) that could help train athletes in learning the complex rowing movements involved in boat racing. The athletes sit in a racing boat that is half physical and half virtual. Projectors are used to create a river scene in front of the athlete and robots are used to provide augmented feedback while the athlete is rowing. Mixed-reality is also often used in the field of architectural design and interior design where 3D models of a building and their interior are created by iterative fine-tuning.

Mixed-reality team meetings are supported by the Holoport project (Kuechler and Kunz, 2006) where a physical table with participants is complemented by a virtual table with participants at a distance giving the illusion that all team members are collocated around a single table.
Fig. 2.10. Mixed reality used as a rowing simulator (left) and in designing buildings (right) (Costanza et al., 2009).

Fig. 2.11. Mixed reality used within the Holoport project (Kuechler and Kunz, 2006).

The prototypes presented above should provide an initial glimpse of mixed-reality environments or physical-virtual environments (Pederson, 2003) that are useful in understanding concepts like physical-virtual artefacts, physical-virtual situations and physical-virtual activities (refer to chapter 3). The following three sections describe existing work in wearable computing, smart objects and smart environments. The survey to be presented is useful in understanding the existing work in the field and such works are inspirational in designing of the easy ADL ecology. The material provided in this survey could be basic knowledge for a reader with advanced
knowledge in ambient intelligence, but for a wider audience this survey will be useful as a starting point in understanding the existing infrastructures for ambient intelligence.

2.4 Wearable Computing

Wearable computing is realized by augmenting a human agent. Wearable computers are small portable computers that are designed to be worn on the body during their use. They are expected to be invisible during usage, should not unnecessarily demand the wearer’s attention, but interrupt the wearer if and when necessary. They should also enable the wearer to perform their physical activities without any hindrance. In that sense, PDAs are usually for hand-held use and are not considered as wearable computers. Wearable computers are usually either integrated into the user’s clothing or into everyday objects that are constantly worn like a wrist watch, belt, shoes, eye glasses, pendant, tie, cap, etc. The aim of wearable computers is to be a constant companion to the wearer and to provide ubiquitous information access and control in an unobtrusive manner.

2.4.1 Definitions of Wearable Computing

According to Licklider’s vision of Man-Computer Symbiosis (Licklider, 1960), “The hope is that, in not too many years, human brains and computing machines will be coupled together very tightly and that the resulting partnership will think as no human brain has ever thought and process data in a way not approached by the information-handling machines we know today”. To achieve Licklider’s vision, computers should be thoroughly integrated into a human agent’s life, being a constant companion with the ability to extract information from the human agent’s environment and gracefully facilitate the human agent’s interaction with the environment.

According to (Starner, 2001), “wearable computing pursues an interface ideal of a continuously worn, intelligent assistant that augments memory, intellect, creativity, communication, and physical senses and abilities.” A wearable computer should have several key attributes including the ability to provide access to information services anytime, anywhere, to be context-aware and adapt a human agent's interaction with computing systems based on their context, and to augment and mediate interaction with the wearer’s physical environment as well.

2.4.2 Wearable Computing as an Infrastructure for AmI

Wearable computing possesses features that are useful in moving towards the vision of AmI. AmI usually focuses on augmenting a human agent’s environment, while wearable computing focuses on augmenting a human
agent and the clothes worn by them. Wearable computing could complement some of the existing research efforts within smart environments. Especially since wearable sensors are closer to a human agent and in many cases are attached to their body, they could be useful in obtaining information about the human agent’s body posture, their limb and other body part movements, physiological states (like blood pressure, body temperature, blood oxygen, electrocardiography, etc.), hand gestures and manipulations, etc. that are usually hard to sense using environmental sensors. Since human agents are mobile agents, it would be hard to imagine environmental sensors that would scale up to all the various locations visited by individual human agents. Also, since wearable computers are personal devices, and are expected to be continuously worn, they could allow smart environments to introduce a sense of personalization and agent-centricity.

2.4.3 Head-Mounted Display and QBIC Belt Computer

Head-mounted displays are commonly used within wearable computing where a human agent’s one eye is covered by a small screen that presents virtual information while leaving the other eye free for interaction in the physical world. EyeTap (Mann, 2004) is a personal experience capturing system that includes both a camera and a display augmented on a human agent’s eyeglasses. It captures images exactly similar to the visual perception of the human agent wearing the EyeTap thereby creating a virtual account of the physical experiences that could be played back or further analyzed in the future. QBIC wearable computing platform (Amft et al., 2004) is a wearability driven design approach towards a wearable computer integrated within a fully functional belt. The main electronics are embedded inside the buckle of the belt, and the belt is used as an extension bus and for mechanically supporting add-ons. QBIC belt computer runs GNU/Linux operating system and is based on Intel XScale family processor. It supports Bluetooth, Radio Frequency, RS-232 and USB communication. Its ergonomic and mechanical design makes it favorable for wearing it in everyday environments.

2.4.4 Gesture Pendant and “Sixth Sense” Computer

Gesture pendant (Starner et al., 2000) is a small pendant that contains a wireless camera and can be worn around the neck. It allows ordinary household devices for instance home theater system, lighting, kitchen sink, etc. to be controlled with hand gestures that are considered natural for human agents as shown in Fig. 2.13 (first from left). Gestures are performed in front of the pendant. Hard to use and understand remote controls can be replaced by simple hand gestures. The “sixth sense” computer (Mistry and Maes, 2009) is a wearable gesture interface comprising of a pocket projector, a mirror and a camera. The projector and the camera are connected to a mobile computing device in the human agent’s pocket. The projector projects visual
information on walls and other physical objects while the camera captures the human agent’s gestures with computer-vision techniques.

![Image: Common parts of a wearable computer. Head-mounted display in the top row and QBIC integrated belt computer in the bottom row (Amft et al., 2004).]

**Fig. 2.12.** Common parts of a wearable computer. Head-mounted display in the top row and QBIC integrated belt computer in the bottom row (Amft et al., 2004).

![Image: Gesture-based interaction with wearable computers. Gesture pendant to the left (Starner et al., 2000) and “Sixth sense” computer to the right (Mistry and Maes, 2009).]

**Fig. 2.13.** Gesture-based interaction with wearable computers. Gesture pendant to the left (Starner et al., 2000) and “Sixth sense” computer to the right (Mistry and Maes, 2009).

### 2.4.5 Electronic Textiles

Electronic textiles are becoming important for wearable computing: the textile fabric is conductive and is used for supporting wearable sensors, actuators
and interfaces. Smart textile is a term that is used for electronic textiles that can sense and react to environmental stimulus in an intelligent and adaptive manner. Fabric-based construction kits like Arduino LilyPad Kit (Buechley et al., 2008) and EduWear (Katterfeldt et al., 2009) have enabled the exploration of future of fabrics and electronic crafts.

![Fig. 2.14. Arduino LilyPad (Buechley et al., 2008).](image)

### 2.4.6 Power Harnessing

Wearable computers require power for operations, and a power conscious design alone is not sufficient to address battery issues. It is not to expect the wearer to keep replacing batteries periodically. Mechanisms to unobtrusively charge the wearable computers are required and are researched in (Kymissis et al., 1998). Human agents dissipate energy into the environment through their actions like walking. Instead of forcing human agents to perform deliberate actions, this work attempts to make use of their everyday actions in generating power. Sport sneakers with an energy dissipating sole made of piezoelectric material or a rotary magnetic generator are used to harness power through bending of the sole, high pressure exerted in a heel strike, etc. Refer to fig. 2.20. Shoes built using piezoelectric sol that periodically broadcast a digital RFID as the wearer walks described in (Kymissis et al., 1998).

### 2.4.7 Personal Server

Intel’s Personal Server (Want et al., 2002) is a mobile device that allows a human agent to store and access their data and applications through interfaces available in the human agent’s environment. Mobile devices in general offer relatively poor user interfaces due to their limited size. Instead of polishing mobile user interfaces, the concept with personal server is to wirelessly connect to input/output devices found nearby like public displays, wireless keyboards, etc. that are not limited in their interaction capabilities,
yet allow a human agent to be mobile and utilize computing power. The personal server is not restricted to mobile context and could be considered as an important constituent in realizing ambient intelligence.

![Image of power harvesting shoes and Intel's Personal Server](image)

**Fig. 2.15.** Power harvesting shoes (left) (Kymissis et al., 1998) and Intel's Personal Server (right) (Want et al., 2002).

### 2.4.8 Body Area Networks

Body area networks (BANs) refer to wireless networking of wearable devices worn by a human agent. BANs are often end in fitness, sports, military and entertainment applications. DexterNet (Kuryloski et al., 2009) is an open-source platform for developing wireless body sensor networks that are operable in both indoor and outdoor environments. DexterNet supports heterogeneous body sensors, and provides an on-node signal processing library called SPINE. DexterNet is used in the Avatar application where a network of motion sensors are used to reconstruct the wearer’s body motion in real time and the visualization of the wearer’s body movements is used for facilitating tele-healthcare. The body motion information could also be used for activity and/or action recognition. DexterNet is also used in the field of public health to understand the association of human exposure to environmental hazards like air pollution, etc. to relevant diseases. In (Yoo et al., 2008), body channel communication is studied where a human agent’s body is used as a communication medium. Comparing body sensor networks to wireless sensor networks in general, body sensor networks are expected to possess fewer and smaller nodes in terms of their size, better energy efficiency to keep the batteries alive for long periods of time, and to allow the wearer to wear comfortably in everyday context.

### 2.4.9 Commercial Wearables

Motorola and Frog Design’s “Offspring” concept includes a set of wearable devices that communicate through Bluetooth forming a personal area network. The wearable digital assistant (WDA) as shown in Fig. 2.16 (left) is
a central device that acts as the hub, while the other devices (Fig. 2.16 right) are useful for specific purposes like the goggles that contain a heads-up display, camera, ear phones and a microphone. The aim is to have these wearable devices function as normal goggles or a wrist watch, but with additional virtual functionalities. At present, “Offspring” is just a concept, but still could be considered interesting in moving towards the future of commercial wearable devices.

![Fig. 2.16. Motorola and Frog Design’s “Offspring” concept.](image)

### 2.5 Smart Objects

Smart objects are realized by augmenting everyday objects. Smart objects are everyday physical objects that provide virtual services without compromising their primary established purpose. Virtual services include sensing, inferring context, communicating information with other smart objects, actuation, interacting with human agents, etc. In providing virtual services, smart objects are expected not to compromise on their appearance and their interaction metaphor. Smart objects fall within a wide spectrum with varied computational and technological capabilities. They range from passively tagged everyday object where the so-called smart object’s intelligence is present in an external infrastructure to everyday objects with built-in intelligence that could act autonomously like intelligent robots. Smart objects are expected to assist and support human agents with their activities without causing any hindrance. According to (Kawsar et al., 2008), smart objects could be classified into three categories namely: standalone and self-contained smart objects; standalone and cooperative smart objects; and backend infrastructure dependent smart objects. Smart objects are usually augmented with sensors that enable setting up wireless sensor networks. Wireless sensor networks are often used for monitoring purposes.
2.5.1 Definitions of Smart Objects

According to (Beigl et al., 2001), “an everyday artifact is defined as a non-computational physical entity with established purpose, appearance and use in everyday experience, while a digital artifact is defined as an everyday artifact augmented with computing and communication enabling it to establish and exchange information about itself with other digital artifacts and/or computer applications.” The term digital artifact is similar to, if not the same as the notion of smart object used in this thesis. According to this definition, smart objects are ordinary physical objects with additional capabilities of computing and communication which enable smart objects to speak with other smart objects, exchange and share information about themselves and their context and become a part of computing applications. Augmenting physical objects with virtual functionalities that matches the object’s primary purpose is an interesting and largely unexplored challenge. Aesthetics and other artistic factors that are not technical by any means come into play in designing smart objects. Augmentation platforms that make it easy to design, build and evaluate smart objects in an iterative manner are also required.

According to (Kawsar et al., 2008), a smart object is a “computationally augmented tangible object with an established purpose that is aware of its operational situations and capable of providing supplementary services without compromising its original appearance and interaction metaphor. Supplementary services typically include sharing object’s situational awareness and state of use; supporting proactive and reactive information delivery, actuation and state transition.” According to this definition, smart objects should be context-aware, i.e. capable of understanding its operational situations in providing support to human agents. This definition of smart object also stresses the importance of not compromising on the original functionalities and usage of the ordinary physical object that is augmented with computing technology.

2.5.2 Smart Objects as an Infrastructure for AmI

Smart objects possess features that are useful in moving towards the vision of AmI. Human environments are usually populated with physical objects. Replacing (some) plain physical objects with smart objects provides the infrastructural platform necessary in making the environment as a whole behave in a “smart” manner, usually researched in smart environments community. Sensors embedded in smart objects are useful in obtaining information about the objects’ states and state changes, surrounding contextual information, activity context in which the individual objects exist, human agents’ interaction with individual objects, etc., and the actuators embedded in smart objects are useful in facilitating remote actuation, automation, information presentation, etc. Smart objects allow for sharing of resources among multiple human agents by making smart decisions and introduce a sense of object-centricity useful within smart environments occupied by multiple human agents.
2.5.3 Media Cup

Teco’s media cup (Beigl et al., 2001) is an ordinary coffee cup augmented with sensors, processing and communication in its base such that the augmentation is invisible and the coffee cup supports context-awareness. Media cup presents its state information like “cup is stationary”, “cup is moving”, “drinking out of the cup”, and “fiddling with the cup”. It also presents additional state information like “current temperature”, “filled up”, “cooled off”, etc. Coffee cups in general have several states and can be used in varied situations making it an interesting artifact to research with. Media cup communicates with other digital artifacts in the environment through infrared communication. Media cup is developed based on an artifact-centered design and is considered to be a standalone and self-contained smart object.

Fig. 2.17. Examples of smart objects: Media cup to the left (Beigl et al., 2001), and AwareMirror (sentient artifact) to the right (Fujinami et al., 2005).

2.5.4 Sentient Artifacts and AwareMirror

Sentient Artifacts (Kawsar et al., 2005) (another term for smart objects) are everyday objects augmented with technology to provide value added services in addition to their primary purpose. Sentient artifacts support interactions that are implicitly driven by sensors and in parallel support explicit interactions that are natural to access those value added services. They are also not restricted to physical artifacts, in the sense that a scheduler or a weather forecast monitor could be considered as virtual sentient artifacts. Prottoy (Kawsar et al., 2008) is a middleware developed for application developers to make use of sentient artifacts in their applications. Prottoy is a generic interface to sentient artifacts making it easy for the application developers to build applications without having to deal with the underlying details about the environment, and in a rapid manner. Virtual artifact
encapsulates the smart object’s environment and allows the application developers to interact with the actual physical artifact with ease.

AwareMirror (Fujinami et al., 2005) is an ordinary bathroom mirror that is augmented with technology to be used as a personalized display. AwareMirror presents the image of the human agent in front of it, but in parallel displays context-relevant information to the human agent. Identifying the human agent in front of it is an important problem which is solved by associating the tooth brush used by the human agent to know their identity information. Proximity of the human agent in front of the mirror is also considered in presenting information. AwareMirror is constructed using an acrylic magic mirror that allows bright colors to seep through in parallel to reflecting images, and an ordinary computer monitor to generate bright images. AwareMirror is considered to be a sentient artifact (Kawsar et al., 2005) that preserves its purpose as a bathroom mirror in parallel to offering digital services in an unobtrusive manner. AwareMirror is shown in Fig. 2.17.

![Image of AwareMirror](image.png)

Fig. 2.17. AwareMirror. (Fujinami et al., 2005).

![Image of Cooperative Chemical Containers](image.png)

Fig. 2.18. Cooperative chemical containers (Strohbach et al., 2004).

2.5.5 Co-operative Artefacts

Smart objects that are standalone and self-contained provide valuable digital services, however there are situations when there is a need for them to communicate and cooperate with each other. Such cooperation could yield a more powerful smart objects system. Cooperative artefacts (Strohbach et al., 2004) is one such system where the chemical containers cooperatively assess their situation in the environment without the need of an external infrastructure in the environment by communicating and cooperating with each other. The smart chemical containers contain embedded domain
knowledge and sensors, and use a rule-based inference to enable cross-artifact reasoning and collaborative inference of knowledge that is hard for an individual smart chemical container to infer in isolation. Cooperative artefacts would function in a wide range of environments since it is not dependent on any backend infrastructure for computing. Refer to Fig. 2.18.

2.5.6 Smart Pizza Packing

The Smart Pizza Packing (Schneider and Kroner, 2008) is an ordinary pizza packing with the additional functionalities of being able to store information about the environmental conditions in which the packing took place by communicating with external sources like sensors and processors. The digital object memory in the pizza packing can be used by applications in the environment or by the human agent in many different ways. The digital object memory is part of the pizza’s whole lifecycle from the pizza production factory to the shipment logistics container to the retailer store and finally to the consumer’s fridge. The smart pizza packing is backend infrastructure dependent where the packing itself contains only a RFID tag, while RFID readers and antennas are instrumented in the environment to make use of the tag information and to store further information about the environmental conditions into the tag. Such an approach was taken since the smart pizza packing would be too expensive even in the near future with processing within the packing itself.

![Smart Fridge](image)

Fig. 2.19. A smart fridge equipped with a touch screen and check surface contents via RFID readers and antennas. The smart pizza packing is intended to be placed within such smart fridges at the consumer’s place (Schneider and Kroner, 2008).
2.5.7 Smart Floor

Smart objects are not restricted to augmented physical objects, but might also include architectural structures like floor, walls, ceiling, etc. Smart floor (Orr and Abowd, 2000) is a system used for recognizing the identity of people based on their footstep force profiles. Smart floor is made of ordinary tiles augmented with load sensors. Smart floor could be used for instance in a home environment (like in the aware home (Kidd et al., 1999)) or in a working environment even when the environment is dark or noisy. Smart floor could be used as a complementary technology to other means of identifying human agents in an environment.

2.5.8 booTable

The booTable (Grammenos et al., 2010) is an interactive coffee table: information is projected on the table’s flat surface. The booTable is furniture that provides additional services like book and DVD information retrieval, photo album and slide show presentation, support virtual chess game, simulates a DJ console, support family notes, shows a virtual table clock, etc. Fig. 2.20 shows a chess game being played on the booTable. A tabl-e-cloth representing an empty chess board is placed on the booTable. The chess coins are projected on the virtual board and the players could either use their IR fingertips or pens to play the game of chess. booTable is also used for fuzzy painting where abstract illustrations can be drawn using the IR finger tips or pens that represent a painting brush. Fig. 2.20 shows a photo album on the booTable with the possibility to navigate and select photo albums using physical hand manipulations. booTable presents a slideshow of photos depending on the table’s state. There are features to add pictures from a mobile phone by simply placing the mobile phone on the table. The mobile phone is RFID tagged, and the pictures are transferred using Bluetooth.

Fig. 2.20. booTable used as a chess board with virtual chess coins (left) and is used as a photo album (right) (Grammenos et al., 2010).
2.5.9 Commercial Smart Objects

Ambient devices (AmbientDevices, 2011) have developed many commercial everyday artifacts that are augmented with digital technology. Examples include an alarm clock that contains weather forecast information which gets updated by connecting to the internet, an ambient umbrella that glows if there is a rain forecast so that the human agent does not forget to take it, an ambient Orb made up of frosted-glass that glows in different colors to display the real-time stock market trends, and a refrigerator from LG Electronics that provides weather forecasts and other useful information to human agents. Refer to Fig. 2.21.

![Fig. 2.21. Ambient umbrella (left), ambient orb (center) and LG refrigerator (right) (AmbientDevices, 2011).](image)

2.5.10 Interaction with Smart Objects

Facilitating a human agent's interaction with smart objects is yet another challenge in designing smart objects. Smart objects are usually tangible and allow human agents to interact with them and obtain tactile feedback. However, there is a thin line between providing additional virtual services and usability. In some cases, it is important to present the virtual services so that the human agent can decide whether to proceed further with those services or not. Some smart objects might be limited in their output modalities to conserve their physical appearances, and it is impractical to provide all the virtual services as tactile feedbacks. In (Kawsar et al., 2010), two interaction techniques are proposed. Refer to Fig. 2.25. One is using a magic lens projection where contextual information is presented on a mobile device screen while the other is a personal projection technique where the information is projected onto the nearest surface with a larger area using mobile projector attached to the mobile device. Smart objects are labeled with 2D barcode which acts as an identity cue to be recognized by the mobile camera.
Fig. 2.22. Smart object specific information and service interface. Presented on the phone screen once an object is tracked in the camera preview screen (top) and projected onto a nearby flat surface, the phone screen being used as track pad (bottom) (Kawsar et al., 2010).

Fig. 2.23. Energy aware kettle (top-left), smart medicine box (top-right) and smart books (bottom) (Kawsar et al., 2010).
Refer to Fig. 2.26. Smart kettle (Kawsar et al., 2010) is an ordinary kettle augmented with software components that provides energy consumption information. It also estimates the energy consumption on a daily and monthly basis. Smart medicine box is intended for medication management where information about the medicine in context and the location of it in the cabinet are presented. Smart book is an ordinary book with additional digital features like presenting reviews of the book from popular online bookstores.

The surfaces of the smart objects are augmented with virtual information using projector-camera system in (Molyneaux et al., 2007). The smart objects are located, tracked and virtual information presented using a projector-camera system that is intelligently computer controlled to move in two dimensions – horizontal (pan) and vertical (tilt). The virtual image projected on the smart object considering a suitable surface on the smart object, where the image is geometrically corrected, and is undistorted. Keep track of smart objects entering and leaving a room is detected. Changes to smart object’s appearance due to changes in intensity, color and direction of lighting is addressed. The system is deployed in a warehouse scenario.

Fig. 2.24. Warning message projection on two chemical containers (left). Scale and rotation invariant local features detected on chemical containers (right) (Molyneaux et al., 2007).

2.6 Smart Environments

An environment populated with wearable computers and smart objects provide the foundation for a smart environment. However, a smart environment is much more complex than just a collection of wearable computers and smart objects. The smart environments research community addresses challenges at the level of considering the environment as a whole, instead of focusing on specific devices or a set of devices usually dealt with in the smart objects and the wearable computing research. Smart environments usually contain plain everyday objects, smart objects, electronic appliances, furniture, devices, autonomous robots and even architectural structures, which are seamlessly interconnected with an overall aim of providing support to human agents within such environments. Smart environments are expected to make use of the human agents’ abilities and compensate for their
limitations in a calm and ambient manner without overwhelming them. In short, smart environments should improve the human agents’ experience.

2.6.1 Definitions of Smart Environments

There are several definitions that describe what smart environments are. According to the EasyLiving project (Brumitt et al., 2000), a smart environment (also referred to as “intelligent environment”) is “a space that contains myriad devices (stationary or mobile) that work together to provide users access to information and services. A broader goal is to allow typical PC-focused activities to move off a fixed desktop into the environment as a whole.” This definition of a smart environment is restricted to the human agents’ virtual activities, i.e. activities that human agents perform in the virtual world or computer world (Pederson, 2003). The activities that were once performed using desktop computers are now researched in an attempt to move them into the human agents’ physical environment. Such a definition means that the physical activities of the human agents are not considered to play any important role.

According to (Coen, 1998), smart environments are “highly embedded and interactive spaces intended to bring computation into the real, physical world. Such environments allow computers to participate in activities that have never previously involved computation and to allow people to interact with computational systems the way they would with other people: via gesture, voice, movement and context.” This definition mentions the fact that smart environments are supposed to not only support activities that are previously supported by computers (typically office activities in the virtual world) but also to support activities that were not previously supported by computers like everyday physical activities at home, monitoring a vineyard, assistance to newly employed trainee in a manufacturing plant, etc. Introducing computers to activities that were previously performed without their assistance would alter the way human agents perform such activities. It is also more likely that such activities shift from being purely physical or virtual to being more physical-virtual as described in (Pederson, 2003). This definition stresses the importance of facilitating human agents to interact with computers in a natural way through multiple modalities.

According to (Youngblood et al., 2005), a “smart environment is one that is able to acquire and apply knowledge about its environment and to adapt to its occupants in order to improve their experience in that environment.” According to this definition, smart environments are expected to improve the human experience within it. This could be done by making the environment adaptive to the human agents by sensing and aggregating knowledge about the environment and its occupants. The term “environment” is usually used within the smart environments community with a bias towards the physical environment; however, the term environment could also be used to refer to the virtual environment in which human agents perform everyday virtual activities.
2.6.2 Smart Environments as Infrastructure for AmI

Smart environments and ambient intelligence are closely related with a typical focus on everyday life of human agents and the possibilities to make it better with technology. While ambient intelligence is more visionary, bringing in techniques and technologies useful for a post-desktop model of computing, smart environments are more confined to specific environment or a collection of environments. Ambient intelligence could be realized through smart environments as an infrastructure, as much as it could be realized through smart objects and wearable computers. In this thesis, the attempt is made to design, implement and evaluate a smart environment, more precisely a smart home, as a means of realizing and exploring the visions of ambient intelligence.

2.6.3 Smart Homes

Smart environments include private spaces like smart homes, smart classrooms, smart cars, etc. and public spaces like smart airports, smart streets, smart hospitals, smart shopping malls, smart trains, etc. Designing systems for smart private spaces and smart public spaces have inherent differences; however, there are many challenges that are common for both types of spaces. In this thesis, the attempt is made to solve some of the common challenges, specifically ignoring the differences for simplicity. Smart homes are a typical example of smart environments. Human agents spend a considerable amount of time at home. This has motivated several research projects to focus on smart homes. Many of the smart homes to be discussed later have a common goal to improve the comfort and productivity of its occupants by predicting the occupants’ immediate needs. What would be the occupants’ immediate need(s) is an interesting and often difficult question to answer. The occupants’ needs vary among home automation, minimizing the home operational cost, obtaining support for everyday home activities, maintaining safety and security at home, providing home entertainment, monitoring the health status of its occupants, automated online shopping, facilitating social interaction, and more. Instead of focusing on the applications for a smart home, one approach would be to investigate a general-purpose infrastructure for smart environments as described in chapter 4.

2.6.4 MavHome and Neural Network House

Smart homes are researched with varying themes. Home automation by observing the occupants and learning to predict their needs is a central theme for much work including the MavHome (Cook et al., 2003) and the Neural Network House (Mozer, 1998). MavHome (Managing An Intelligent Versatile Home) is an agent-based approach with an aim of maximizing the occupant’s comfort and minimizing the operational cost. The MavHome architecture is
made up of a hierarchy of multiple agents that cooperate and coordinate in an effort to achieve the goals of the MavHome. The MavHome not only focuses on automating devices, but also predicts the occupant’s activities and the actions within it, and uses this information in deciding whether to automate or not. The SHIP (smart home inhabitant prediction) algorithm is used for action prediction: the inhabitant’s interaction with the devices is recorded and the sequence history is used to predict the next possible action. The Neural Network House (Mozer, 1998), similar to the MavHome, attempts to program the home based on the lifestyles and desires of its occupants that are learned through observation. The neural network house learns to anticipate and accommodate the occupant’s needs with respect to basic comforts like air heating, lighting, ventilation, and water heating. (Liang, 2002) is another agent-based approach to home automation. Commercial home automation technologies like Lonworks and X10 offer networking, automation and remote access possibilities, but offer limited support for learning and predicting a human agent’s needs within a smart environment.

Fig. 2.25. Automated blinds in the MavHome (Cook et al., 2003).

2.6.5 iDorm

The iDorm (Holmes et al., 2002) contains everyday furniture and physical objects that are usually found in a dormitory. However, many of the furniture and physical objects are fitted with sensors and effectors: pressure sensor on a chair, pressure sensor on a bed, occupancy sensor, telephone status sensor, internal temperature sensor, humidity sensor, internal light level sensor, etc. are embedded along with effectors to control window blinds, desk lamp, PC-based word processing application, PC-based media playing application, cooler fan, heater, dimmable spot lights, etc. The iDorm supports multiple human agents within it and the human agents could control the devices within it through a PC interface, an iPAQ interface, a mobile phone interface and an iFridge (internet fridge from LG electronics) interface with a touch screen capabilities as shown in Fig. 2.26. The iDorm supports three
Emerging Interaction Paradigms and Approaches

heterogeneous networks: Lonworks, 1-wire (TINI), and IP. The iDorm (Doctor et al., 2005) reasons about its occupants using fuzzy techniques and uses such an inference to satisfy both the occupant and the system objectives. Occupant behaviors and the environmental conditions are modeled to be adaptive in the sense that the iDorm allows for modifying the learned behavior online and using a life-long learning mode.

![Fig. 2.26. iDorm interfaces: a) PC interface, b) iPAQ interface, c) mobile phone interface, d) iFridge interface (Doctor et al., 2005).](image)

2.6.6 Gator Tech Smart House

The Gator Tech Smart House (Helal et al., 2005) focuses on developing technologies that are scalable, cost-effective and extensible such that they evolve to accommodate newer technologies or further developments in the application domain. Many of the existing smart environments are not extensible and the environments are restricted to original implementers. The Gator Tech smart home middleware is generic and it is available both as a runtime environment and as a software library for third party implementers. The service definitions of for instance the sensors, actuators, etc. are included in the library and the application developers can assemble the services into composite applications. Many physical objects within the Gator Tech smart home are augmented with technology including the mailbox, front door, window blinds, bed, closet, laundry, mirror, bathroom, refrigerator, floor, plugs, microwave oven, displays, etc.

2.6.7 Aware Home

The Aware Home initiated research efforts in investigating issues surrounding computing in a home by serving as a living laboratory for ubiquitous computing research in support of everyday living at home (Kidd et al., 1999). The aware home supports context awareness and ubiquitous sensing, human interaction with the home, an application to find lost objects, and elderly care. Social issues are studied in the aware home. The Aware Home is capable of capturing information about itself, its occupants, and their activities. Capturing the occupants’ activities and needs is challenging to achieve in practice, which has led to approaches that are semi-automated,
thereby giving the occupants more control as in the House_n project (Stephen, 2002).

Activity recognition performed in a living laboratory setup might not give results that generalize to real situations. Human agents occupy such smart environments for a short period of time, might not react naturally as they would for instance in their home environment, might be influenced by the smart environment and what is available in it, instead of what they would like to have, and overall the results could be sensitive to the smart home architecture. Nevertheless, such results and living laboratory experiments are important during the prototype development stage for exploring novel concepts and to improve the current prototypes. In this thesis, the easy ADL ecology to be described in chapter 4 was developed as a living laboratory environment to sharpen and evaluate the concepts, and in parallel to moving towards a working smart home in the physical world using available technologies.

2.6.8 Nursebot Project, SmartBo, Gloucester House and CareLab

Another central theme is to design smart homes that enable elderly persons to live a longer and more independent life at home (in accordance with the concepts of successful ageing and assisted cognition). EU research programs like the Ambient Assisted Living (AAL, 2011) are moving in this direction. Since life expectancy is increasing and more elderly people would populate the society in the future, efforts like Ambient Assisted Living are necessary to provide a good quality life for the elderly. Elderly people are also prone to physical and cognitive disabilities that should be considered in designing smart homes. Independent living means that the elderly person must be able to do basic activities of daily living like taking a shower, eating, dressing, etc. and more complex activities of daily living like preparing a shopping list, preparing food, etc.

Activity recognition plays an important role in an effort to enable independent living of elderly within their smart homes. The Nursebot project (M.E. Pollack et al., 2002) is aimed at developing mobile robotic assistants that support elders in performing their activities of daily living (ADL) at home. Smart homes for people with disabilities are an objective of many research efforts. The SmartBo project (G. Elger, 1998) focuses on elders with mobility impairments and cognitive disabilities while the Gloucester Smart House (Adlam and Orpwood, 2004) focuses on people with dementia. An automated hand-washing assistant for people with dementia is researched in (Hoey et al., 2010). Automated health monitoring and anomaly detection for cognitively and physically challenged people using machine learning algorithms that can model their behavior is described in (Jakkula and Cook, 2008). Ambient assisted living with the CareLab (Ruyter and Pelgrim, 2007) have focused on monitoring and coaching elderly people to enable them to maintain an independent lifestyle by focusing on their feeling of safety, cognitive prosthesis and social interaction.
2.6.9 Typical Challenges within Smart Homes

There are many challenges to address within a smart home context. Four important challenges namely: (a) networking of smart objects; (b) indoor location and object tracking; (c) activity and behavior recognition; and (d) facilitating human-environment interaction has been inspirational in conceptualizing and developing the personal activity-centric middleware components to be described in chapter 4.

2.6.9.1 Networking and the Internet of Things

Networking of smart objects is an important challenge explored by the smart objects and smart environments research community. Smart-Its (Beigl and Gellersen, 2003) is a hardware and software platform that enables everyday objects to be augmented with small embedded sensing, actuation, processing and communication modules that speak to one another wirelessly forming a short-range ad-hoc network. Smart-Its motes are small in size and would not be physically visible when augmented on everyday objects as shown in Fig. 2.28. Some Smart-Its can be very small with a dimension of 1000 mm\(^3\) including battery package and a variety of sensors (Decker et al., 2005). Other Smart-Its modules are slightly bigger, include additional capabilities like an integrated Bluetooth transceiver, or user interfaces.

![Smart-Its motes (Beigl and Gellersen, 2003).](image)

The Internet of Things (Gershenfeld et al., 2004) is about extending and bringing internet to the physical world through physical objects that are connected to the virtual world with an IP address thereby offering additional functionalities. The internet of things is related to smart objects in the sense that smart objects could be used as building blocks for the internet of things as discussed in (Kortuem et al., 2010). RFID is one technology that could be
used to build the internet of things as with the RFID ecosystem built at the University of Washington (Welbourne et al., 2009).

According to (Mattern and Floerkemeier, 2010), everyday objects should possess the following capabilities in order to move towards the internet of things. The objects should be able to communicate and cooperate with other heterogeneous objects and services, and follow open standards. The objects should be virtually located and addressed globally from remote locations. They should include a unique identifier like a RFID tag or visual tag for associating virtual information to the physical object. Virtual information may include its properties, states, ownership, manufacturer information, etc. They should include sensing and actuation capabilities, and use such capabilities to infer context. They should possess embedded processing capabilities, aware of physical location or at least could be physically located, and possess interfaces that are natural for human agents to interact with it.

Fig. 2.28. The person tracker displays the locations and identities of tracked people in a room. The cameras’ fields of view intersect the ground plane and the locations of the people-shaped blob clusters reported from the two stereo modules used for location tracking (Krumm et al., 2000).

2.6.9.2 Indoor Location Tracking

Indoor location tracking of human agents and objects is another challenge within smart homes. Computer vision based identification and tracking of human agents within a smart living room is described in (Krumm et al., 2000) as part of the EasyLiving project. Location and identity of human agents are important contextual information for facilitating smart home services in real-time like triggering events based on agent location, locating the right device to present information, invoking personalized ambience in the room occupied by a specific agent, understanding an agent’s behavior to assist them, etc. Two sets of color stereo cameras are used that track multiple human agents
sitting, standing, walking, and entering and leaving the living room. The two-
dimensional ground plane location information of the agents is used by a
person tracker application that simulates a map of the living room as shown
in Fig. 2.28.

The Active Badge system (Want et al., 1992) tracks objects in an
environment using badges that are attached to individual objects. An active
badge periodically transmits its unique identity information using infrared to
receivers that are placed in fixed locations in an indoor environment. Indoor
walls act as a natural boundary, enabling the active badge to be tracked
within a room. The Active Bat system (Hightower and Borriello, 2001)
developed at the AT&T lab uses ultrasound time-of-flight for location
tracking. Human agents and objects are augmented with active bat tags
which are small wireless transmitters. A tag sends ultrasound signals with
identity information to a matrix of ceiling mounted receiver. A central radio-
frequency base station calculates the distance of the tag with reference to
several receivers, and determines the location of the active bat tag. The
Cricket location support system (Priyantha et al., 2000) uses ultrasound
emitters in the environment and ultrasound receivers in the objects to enable
localized location computation on individual objects. Radio-frequency signals
are used for synchronization and for removing reflected and indirect
ultrasound signals.

RADAR (Bahl and Padmanabhan, 2000) is a WLAN-signal-strength-
based location-tracking system that attempts to make use of existing radio-
frequency infrastructures within smart environments. During a training
phase, a signal-strength map of the environment is built by placing a
transmitter in different locations and measuring its signal strength at
individual fixed receiver locations. During the recognition phase, the
transmitter location is determined based on the received signal-strength
values at the different fixed receiver locations. Triangulation algorithm is
used to determine the transmitter location using received signal strength
indication (RSSI) at three or more receivers which could be an object or a
human agent. A location tracking approach that makes use of a single WLAN
access point is described in (Zárua et al., 2007). Received signal strength
indication (RSSI) from a single radio frequency interface is used to track
location with sub-room level precision. However, for accurate location
tracking, triangulation-based approaches are preferable.

Indoor location-tracking using location fingerprints (a database
containing pre-recorded signal strength measurements) is described in
(Mikkel Baun, 2011). Signal-strength measures at a wireless client are
compared to the existing location fingerprints in determining the client’s
location. An automatic mapping-based method is used which avoids the need
to manually calibrate the signal-strength measures. The smart floor (Orr and
Abowd, 2000) described earlier contains embedded pressure sensors that
capture a human agent’s physical contact with the floor in determining the
human agent’s location, based on pre-recorded walking patterns, the human
agent’s identity, and foot falls. The smart floor removes the need to carry
additional devices or specific tags, but suffers from scalability issues.
However, within a smart home context, scalability should not be a major issue
considering the size of typical homes. Proximity-based location tracking is another approach where the location of objects or human agents is determined based on their proximity to locations already known. RFID technology is typically used in proximity-based location tracking (Song et al., 2007).

2.6.9.3 Human Behavior and Activity Recognition

Recognizing human behavior and the activities performed by occupants of a smart home is another interesting challenge. Multimodal detection of human agents' behavior in a smart home using a 3D video tracking system and scene analysis is described in (Brdiczka et al., 2009b). First the human agents' body postures are recognized using the 3D video tracking system in combination with support vector machines (SVM). Body postures include “standing”, “lying down”, “sitting”, etc. Then body posture information is combined with other contextual information like speed, distances, etc. to determine the human agents' roles. Behaviors include “walking”, “interacting with the table”, etc. Their behaviors in combination with speech activity detection through a headset microphone worn by the human agents and ambient sound detection using microphones embedded on the apartment walls are fed as input to Hidden Markov Models (HMMs) that determine the situation within the smart home. The rich data obtained by observing human agents could be used offline for further analysis as well as online to provide timely support to the occupants depending on the application.

Fig. 2.29. iglove (left) and iBracelet (right) (Smith et al., 2005).

Activity recognition based on single occupants in a home environment is investigated in (Cook and Schmitter-Edgecombe, 2009, Philipose et al., 2004, Lester et al., 2006, Logan et al., 2007, Kasteren et al., 2008). The sequence of objects used by a human agent in performing activities is used for activity recognition in (Philipose et al., 2004). Activities are represented as probabilistic sequences of object usage. Objects are tagged with passive RFID tags that are stamp size, battery free, durable and inexpensive. An RFID reader is worn by the human agent on their hands as a glove (iGlove) with an antenna in his palm as shown in Fig. 2.29. The iGlove detects the objects that are currently being manipulated by a human agent which could be useful for
Emerging Interaction Paradigms and Approaches

recognizing their activities. In the future, RFID-reader-enhanced wrist watch or bracelet could solve the ergonomic issues as with the iBracelet (Smith et al., 2005). iGloves can be used only for certain activities, while an iBracelet could be used for a variety of activities, is more aesthetical and ergonomical. iBracelet retains the features of iGlove from a functional perspective. From the tagging of physical object’s perspective, it is not that all the objects in an environment will be tagged. Only selected and useful objects are tagged. The Electronic product code (Code, 2011) is an initiative to support the use of RFID technology by tagging objects and tools of interest.

Wireless Identification and Sensing Platform (WISP) (Smith et al., 2005) does not contain a battery and are powered by line-of-sight readers. However, WISP contains anti-parallel mercury switches that get connected depending on the tilt of the object. In this way, not only the identity, but also the motion information is obtained in the reader without a power source on the WISP.

(Kasteren et al., 2008) performed activity recognition in in-situ home environments by deploying a sensor network complemented by Hidden Markov Models to recognize the activities performed by the home occupants. An annotation system was also built to enable the occupants to provide the ground truth of the activities that they are currently performing to the activity modeling system. The in-situ experiments give higher credibility to the results in comparison to living laboratory experiments. The sensor data and the annotations are available online to be used for research purpose.

Multi-Agent Environments. Home environments in general are not restricted to single occupants and usually include multiple occupants. Recognizing the activities performed by multiple occupants of a smart home environment is challenging since sensor events could fire for several activities in parallel and to keep track of the multiple activity sequences in a physical world confined to noise and variations in activity patterns is complex. Activity recognition within a smart home comprising of multiple occupants is described in (Wang et al., 2009). The occupants wear a multimodal sensor platform for detecting their hand movements, location, their interaction with objects in terms of touch and sound, their interaction with other human agents through voice and environmental information like temperature, humidity, light, etc. Coupled Hidden Markov Models (CHMMs) are used to recognize both personal activities and group activities. Interestingly, group activities are recognized with better accuracy. This paper argues that coupling the HMMs was useful for group activities, while personal activities performed by individuals were affected by the noise created by the other individual in the environment performing his/her personal activity thereby reducing the recognition accuracy for personal activities. In (Singla et al., 2010), independent and joint activities performed by multiple occupants of a smart home environment are recognized using motion sensors, temperature sensors, item usage, etc. However, many of the activities that were considered are simple activities where a straightforward mapping of the sensor events to the activities could be established.

Assessing the Quality of Human Activities. Instead of focusing on activity recognition alone, (Cook and Schmitter-Edgecombe, 2009) has also focused on
recognizing and assessing the quality with which activities are performed in a smart home that acts as a test bed for further research. The activities performed by the occupants are checked for their completeness, consistency, performance of incorrect steps and specific steps that are skipped. Such a smart home is intended for remote health monitoring and interventions. Five activities, namely telephone use, hand washing, meal preparation, eating and medication use, and cleaning were chosen for the experiments.

2.6.9.4 Interaction within Smart Homes

There are several research efforts (Heinroth et al., 2011, Bellik et al., 2009, Minker et al., 2010) that attempt to explore the possibilities of human-computer interaction within smart environments. According to (Dahl, 2008), the “smartness” of a smart home could be related to the occupant’s situative interaction possibilities. Physical interaction models using location-awareness and token reading were explored (Dahl, 2008), however the virtual counterpart was ignored. Even though many smart homes exist today there have been few attempts to investigate the possibilities of integrating the various physical and the virtual aspects of a smart home. Smart meeting rooms populated with local physical participants and distant virtual participants interacting with each other in a mixed-reality context (mentioned earlier) is a typical example of considering both the physical and the virtual aspects of a smart environment (Nijholt et al., 2009).

Gesture-based interaction within a smart home is described in (Rahman et al., 2009). Human agents wear a glove augmented with infrared emitters, while infrared cameras are used to detect the motion path created by the human agent’s hand gestures to draw different symbols in the air. Gesture-based interaction is usually considered to be natural for human agents and it allows controlling some of the services available in the smart home. Hand gestures are used for turning on and off a lamp switch, to change the light intensity, for pausing and playing a movie and for other security related services. Camera based approaches are sensitive to environmental lighting conditions and this work using infrared technology removes such additional challenges. Such gesture-based explicit interaction could be complemented with implicit interaction using contextual information.

Eye-based interaction for environmental control within a smart home is described in (Corno et al., 2010). Human agents usually touch and manipulate objects in a home environment while this work attempts to provide an alternative solution where in a human agent could control objects using their eye gaze movements. This work is part of the COGAIN (communication by gaze interaction) project funded by the European Commission for human agents with motor disabilities to lead an independent life at home. Human eye gaze could be associated to visual attention and their saccadic eye movements could be used to interact and control objects. However, eye gaze alone does not always represent a human agent’s visual attention described as the Midas touch problem (Jacob, 1991). Three infrared cameras and infrared light sources are used to capture a human agent’s head movements
using the smarteye eye movement system (SmartEye, 2011). Facial features are detected and then the irises and pupils are located to determine the human agent’s gaze direction. The eyelids are finally localized and the eyelid-to-eyelid distance is calculated using a model of the eyeball (Shi et al., 2006). Eye gaze interaction, similar to gesture based interaction is considered to be intuitive and natural to human agents.

### 2.6.9.5 Other Challenges

According to (Edwards and Grinter, 2001), there are many challenges that span across the technical, social and pragmatic domains in realizing the concept of smart homes in reality.

1. **Intelligibility problems:** where the human agents within a smart home are unaware of some of the virtual features of the smart home since they are invisible and inaccessible. Designing artifacts that present not only their physical features, but also their virtual features and services should be incorporated.

2. **Impromptu interoperability:** where the different devices in a smart home are manufactured at different times and by different vendors creating a scenario where the devices have difficulties in communicating and exchanging information with each other thereby hampering the possibility of providing a holistic experience to the human agent within it.

3. **No system administrator:** i.e. occupants of a smart home are usually not system administrators that can fix technical problem that might arise and the smart home should include self-healing mechanisms to enable non-technical human agents to live in it.

4. **Domesticated use:** the objects and devices within a smart home are intended for domestic use. For instance, smart dishes with embedded technology might be in a dishwasher or smart clothes might be in a washing machine. Hence such domestic usage factors should be considered in designing artifacts for a smart home.

5. **Social implications:** smart homes might be occupied by multiple human agents and this introduces social implications like family setup, social routines, etc. that should be considered in augmenting technology within a home.

6. **Handle ambiguous data:** smart homes are dependent on sensor and wireless communication technologies that might be unreliable at times, and inferences in the presence of ambiguity are yet other challenges to be addressed.
Chapter 3

Egocentric Interaction: a Human-Centered Interaction Paradigm

This chapter describes egocentric interaction, a human-centered (both body and mind) interaction paradigm intended to address the visions of ambient intelligence. A brief introduction to the current understanding of human cognition by describing relevant theories related to human cognition, existing cognitive architectures and four important human-related factors facilitating their cognition namely perception, action, attention and intention are presented. Egocentric interaction is conceptually developed based on several basic principles and assumptions like the physical-virtual equity principle, situatedness, the proximity principle, perception and action instead of input and output, etc. A physical-virtual design perspective based on the physical-virtual equity principle is described followed by a situative space model which is an important corner stone for egocentric interaction and is based on the principle of situatedness and the proximity principle. An application of the situative space model to an everyday breakfast situation is described followed by associating the situative space model to human attention and intention. Finally, a discussion and the road ahead in developing egocentric interaction as an interaction paradigm for ambient intelligence are discussed.

3.1 Understanding Human Cognition

Understanding human cognition plays an important role in facilitating computer-mediated human activities, especially for ambient intelligence. This introduces a need to develop models of human cognition for a domain like HCI. Such models are useful at least for two reasons: 1) to utilize the wide range of human cognitive capabilities and to compensate for their limitations; and 2) to create artificial agents that complement human agents.
3.1.1 Theories Related to Human Cognition

3.1.1.1 Situated Cognition

According to the theory of Situated Cognition (Brown et al., 1989), human cognition cannot be separated from context. Here the context could be physical, activity, social, or cultural. Human thoughts and actions occur *in situ*, i.e. adapted to their situation. Human actions to a large extent are improvisatory and are executed by directly coupling perception and action. Situated cognition theory could be considered as a complement to the traditional theories in cognition weighting upon the inferential processes of deliberation and planning. Situated cognition assumes that human knowledge has a dynamic aspect to it, and is formed based on a human agent’s interaction within an environment. Human knowledge is constructed in performing actions instead of being a stored artifact. Human agents are social beings by nature, whose knowledge is created by their active participation in this world and the meaning is created by their ability to actively participate in a meaningful way.

3.1.1.2 Embodied Cognition

Embodied Cognition (Wilson, 2002), similar to the theory of situated cognition emphasizes that cognition is a situated activity that is actively constructed. The importance of environment and its role in the development of cognitive processes is also acknowledged by embodied cognition. According to embodied cognition, embodiment is a necessary condition for cognition, and it could be understood as a human agent’s perception-action capabilities that enable them to interact and survive in an environment. According to (Clark, 1998), human cognition cannot be viewed separately from the body and the surrounding environment. Human agents are embodied beings with a body that constraints the way an environment appear to them and the way they interact in it. An environment would appear differently to different human agents depending on their perspective and their surrounding context. Such a perspective centered personally on individual human agents is referred to as their *egocentric perspective* and their interaction in an environment with such a perspective is referred to as *egocentric interaction*. A human agent’s egocentric perspective is dependent on their current activity, perception and action possibilities, and the environment in which the human agent is situated in. The point that human cognition is grounded on their physical characteristics and their embodied experience in the world differentiates embodied cognition from research in situated cognition, even though the two theories are closely related and complementary in nature.
3.1.1.3 Distributed Cognition

According to the theory of Distributed Cognition (Hollan et al., 2000), cognition is not confined within an individual but is distributed among objects, tools and human agents in an environment. Distributed cognition extends the boundary of analyzing cognition from an individual human agent to the environment as a whole, making it an interesting theory to consider in facilitating HCI within ambient ecologies. Since cognition is distributed, it may be distributed among a group of human agents resulting in emergent social interactions that are important to consider in understanding cognition. It is discussed as collective intelligence in (Lévy, 1997). Similarly, cognition could be distributed among physical objects, resulting in the “offloading” of information onto the environment and thereby creating a medium for social collaboration. Also, such objects in the environment could be seen as an extension of a human agent’s mind, playing an active role in driving the human agent’s cognitive process described by the concept of the extended mind (Clark and Chalmers, 1998). Similar to embodied cognition, distributed cognition supports the embodied nature of cognition, thereby considering the human body and the material environment around them to take a central role in understanding cognition. Human agents live in complex cultural environments resulting in strong relationship between culture and cognition. Human activities in the historic past has shaped human cultures of the present, which influences the way human agents think and act leading to the emergence of future human cultures.

3.1.1.4 Activity Theory

Activity Theory (Kuutti, 1996) is widely used as an analytical and evaluation framework in HCI on aspects relevant to egocentric interaction such as regarding activities as their unit of analysis, regarding the motives and needs of human agents, as well as the psychological and social contexts of their activities. It also shares premises such as shifting from the notion of user to the notion of agency, and the situatedness of human actions. Furthermore, activity theory could provide valuable insights into extending egocentric interaction by including social and cultural aspects, which is not addressed in this chapter and is left for future work. According to Activity Theory (Kuutti, 1996), a human agent’s activity could be modeled using a hierarchy of activity, actions and operations wherein the hierarchy changes dynamically according to the context. Human activities have an objective and could be split into goal directed actions that are achieved by low-level operations.

3.1.1.5 Ecological Psychology

According to ecological psychology (Gibson, 1979), understanding human cognition involves studying a human agent in their everyday environment. Human agents do not exist or act in isolation. Instead, they adapt themselves
to establish a harmonious working relationship with their environment. Ecological psychology emphasizes the influences of a human agent’s situation on their behavior similar to earlier mentioned theories like situated cognition and embodied cognition. Donald Norman (Norman, 1999) is one of a wide range of researchers in psychology and cognitive science that have been influenced by the ecological psychology of (Gibson, 1979), which, of course, has its very focus on the relation between an individual natural agent and the agent’s natural environment. Specifically, Norman has adapted the term affordance from Gibson, using it to refer to the quality of an object that allows a human agent to perceive and perform actions using it. The affordances of an object depend on the human agent’s physical abilities to perceive and act upon it, the agent’s objectives and intentions, past experience in using the object, etc. To egocentric interaction, a situated approach is fundamental: the agent’s abilities to perceive and act are considered from the point of view of how a human agent’s body is situated in the environment, as are the agent’s intentions, objectives and activities.

To summarize, most of the above mentioned theories on understanding human cognition emphasizes the following: 1) the importance of a human agent’s environment; 2) the importance of a human agent’s situation and embodiment; and 3) the importance of viewing a human agent’s environment as a whole including the social and the cultural aspects integrated within it.

3.1.2 Cognitive Architectures

This section describes the existing computational models of cognition for taking inspiration in developing concepts related to egocentric interaction. Cognitive architectures attempt to describe the structure of human cognition in a comprehensive manner. The term was coined by Newell based on his work on computer architecture (Bell and Newell, 1971). Cognitive architectures in general are not developed to address existing problems within HCI, but to establish a wide range of theoretical concepts about human intelligence, their abilities and limitations. Research findings in cognitive psychology usually drive the designing of cognitive architectures. Traditionally, human cognition was viewed to be independent of a human body and its surrounding environment.

3.1.2.1 Model Human Processor

According to the Model Human Processor (Card et al., 1983), the human cognitive processor is composed of three modules namely perception, cognition and action. It views cognitive processing to be a cyclic and sequential process from stimulus perception to cognitive problem solving to response execution. Traditionally, both cognitive psychology and artificial intelligence have focused on the cognitive problem solving and performance aspects of the human agent. Human perceptual and motor systems are complex in their own rights needing appropriate research attention. Also, it is likely that human
cognition is constrained by human perception and motor capabilities, and vice versa. This is in line with the above mentioned cognitive science theories that stress the need to consider human agents as part of a complex system that includes their body and the surrounding environment.

3.1.2.2 SOAR, ACT-R and ICARUS

SOAR (Newell, 1990) and ACT-R (Anderson, 1993) incorporate some aspects of human perception and motor action within their cognitive architecture. ACT-R is based on a serial model of perception, cognition and action, mainly suitable for human interaction with single computer displays and action possibilities that are limited to using a mouse and a keyboard. ICARUS (Langley and Choi, 2006) considers cognition to occur in physical contexts strongly grounded by human perception and action, there by giving importance to physical interaction. ICARUS describes cognitive behavior to occur in recognize-act cycles where a set of perceived objects including their type, names and description are deposited in the perceptual buffer for matching against long-term conceptual definitions before executing actions. ICARUS stores knowledge in the form of concepts that describe environmental situations, and skills that describe the mechanism to achieve goals by decomposing them into sub goals. According to (Kotovsky and Gentner, 1996), human concepts in general are relational in nature and are described by their interactions among objects and events. ICARUS considers such a relational view of objects in describing concepts grounded in perception and skills grounded in executable actions.

3.1.2.3 EPIC, ACT-R/PM, and HiTEC

EPIC (Kieras and Meyer, 1997) is another cognitive architecture that contains a comprehensive set of perceptual-motor mechanisms, but considers the cognitive mechanisms to play a peripheral role. EPIC is based on parallel processing of perception, cognition and action. EPIC is based on a simple production rule system that contains a set of production rules that translate preconditions into actions to attain certain post-conditions. Since EPIC does not handle conflict resolution, it allows all the productions that meet the preconditions to execute actions in parallel which is a limitation. ACT-R/PM (Byrne, 1998) is an extension of ACT-R that considers the perceptual and motor mechanisms to be an integral part of its cognitive architecture, giving them an equal status with the cognitive aspects. ACT-R/PM comprises of features, especially its parallelism that allows for cognition in dynamic environments. ACT-R/PM is composed of a set of perceptual-motor modules that operate in parallel, and can handle situations that demand high performance. ACT-R/PM is restricted to accurately modeling human latency delays. HiTEC (Haazebroek and Hommel, 2009) is a cognitive architecture that includes not only perceptual and action-related information processing but also the interplay between perception and action. This allows for covering some known phenomena in cognitive psychology like the Simon effect (Simon
and Rudell, 1967), that action planning affects the perceptual process (Fagioli et al., 2007) and the perceptual effect of action influences action execution (Hommel and Elsner, 2009) to be discussed in section 3.1.3.

### 3.1.2.4 CogAff and Emile

Human emotions, their attitudes, and motivation play an important role in human actions. The importance of human affective experiences is considered in some cognitive architectures like CogAff (Sloman, 2001) that attempts to support interaction between cognition and affect; and Emile (Gratch, 2000) which considers an agent’s emotions in its action selection, while infer other agents’ emotions in selecting its dialogue choices. To summarize, modern cognitive architectures have acknowledged the importance of a human agent’s perception and action possibilities in modeling human cognition. Further, the interplay between perception and action is also considered important in building a cognitive architecture.

### 3.1.3 Perception and Action

Human perception is a process that enables human agents to interpret and organize sensation to produce meaningful experience of the world (Lindsay and Norman, 1977). Here, sensation refers to the unprocessed result of stimulation of sensory receptors in eyes, ears, nose, tongue, or skin through internal sensors. Many research efforts focus on visual perception for understanding human perception as a whole since 70% of human sensory receptors are visual receptors and vision engages 50% of human cortex. Human action is a process that enables human agents to execute their goals in the world.

As mentioned earlier, human perception and action are closely coupled. The theory of event coding (Hommel et al., 2002) describes the perceptual and the action processes to share a common representational medium with bi-directional interaction between them. The Simon effect (Simon and Rudell, 1967) describe human tendency to react towards the location of the stimulus in performing response action even if the location cue is irrelevant to that action. Action planning can affect perceptual processes, which challenges the traditional view of sequential information flow from stimulus perception to action planning and execution (Fagioli et al., 2007). According to (Fagioli et al., 2007), action affects perception by creating a bias on the objects that are perceivable depending on their relationship to the ongoing or intended action. Object representation and action plan representation overlap to an extent depending on the features that are common between objects and action plans (Hommel et al., 2002). Even in the absence of action execution, subjects scanned using fMRI showed highest activation in the pre-motor areas (areas involved during action) of the brain during the anticipation of perceptual events, thereby indicating it to be an action-relevant stimulus (Schubotz and von Cramon, 2002).
Human agents are capable of learning the perceptual effects of actions and use this knowledge in action execution (Hommel and Elsner, 2009). Further evidence of it was found in a brain-imaging study (Elsner et al., 2002). According to the “ideomotor principle” (James, 1890), human agents build up the desired effects of an action in mind before executing the goal-directed action. This also suggests that actions are goal-directed to an extent instead of considering them as mere responses to stimulus perception. Otherwise, it would also be difficult for a human agent to know if the action performed has actually satisfied their goal or not; and a stimulus-response agent that behaves purely based on perception might not be capable of making “intelligent” decisions which require higher-level cognition. Human agents realize their goals by actually performing actions in varied contextual conditions.

Human actions are not fully pre-specified before the actual execution of the action. Human environment and the object upon which actions are directed play an important role in specifying certain parameters for the execution of an action. According to Gibson’s concept of affordance (Gibson, 1979), an environment or an object affords the possibility of performing certain actions while not allowing for other actions in relation to the actor. Thus, action representations are not completely internal within a human agent and some of the representations are actually distributed in the environment. Apart from the physical effects of action, there is also an affective aspect to action. According to (Thorndike, 1927), actions yielding positive affective consequences are more likely to be executed in comparison to those with negative affective consequences.

For a human agent to achieve their goals by executing actions, the human agent should understand their current situation. Human agents not only perceive objects and events that are part of a situation, but also recognize them. Recognition is closely related to categorization and could be considered as the process of mapping objects, events, etc. to concepts or categories already known to a human agent. Recognition is being able to categorize objects and events into their respective type. Recognition is also closely related to perception in the sense that the output from perception is usually considered as input for recognition, however some cognitive systems view them as the same process. Based on the recognition of objects and events, human agents could potentially examine their situation. The process of examination includes determining the states of objects and events; and in the formation of new concepts by conscious observation and learning.

Recognition and examination of the situation, along with the perceptual effects of action in that situation influences a human agent’s decision to execute an action among alternative actions. Human agents execute action in an environment by selecting the objects that could potentially be acted-upon considering the environmental conditions and the objects’ relevance to their action goals. The selection process is followed by the manipulation process where the objects’ states are modified, even though some cognitive systems might view the two processes alike. The situative space model to be described in section 3.4 includes processes like perception, recognition, examination, action, selection, and manipulation.
According to (Gibson, 1979), human agents do not perceive the entire world of physics and behave in the space and time of physics. They frame the world into an environment where they can survive through perception and action, constrained by their body. Human perception and action capabilities are framed by their body, which is used as a starting point in egocentric interaction. Human agents use multiple modalities to sense and act in an environment. Perception could be considered as an integration of the results from different modalities into a unified description of the environmental situation. Similarly a unified description of an action could be executed through several human modalities.

3.1.4 Attention and Intention

Human agents have limited cognitive processing capabilities for perceiving the world and acting within it. Human attention comes in as a filter that preferentially allocates cognitive processing resources to objects and events that are important for perception and action depending on the human agent’s intentions and the activity context. According to some theories (Duncan and Humphreys, 1989), human visual perception can take place without attention by extracting perceptual primitives from the environment. Experimental evidence of grouping under conditions of inattention was shown in (Moore and Egeth, 1997). However, according to (Mack et al., 1992) there is limited perceptual organization under conditions of inattention.

Divided or peripheral attention in human agents allows them to be aware of stimuli within their visual field without focusing their attention on the stimuli. Note that the same applies to other modalities as well. Such stimuli might be distracting stimuli as in a visual-search scenario, or the stimuli might be secondary task relevant stimuli in a dual-task scenario. Full-attention or central attention in human agents allows them to be aware of stimuli within their visual field by focusing their attention on the stimuli. Such stimuli might be primary task relevant stimuli in a dual-task scenario.

Human agents allocate their attention to a large extent depending on their action context (Yarbus, 1967). For instance, human agents take a top-down approach in search of objects and events relevant to an action context within which they are currently situated in (Hopfinger et al., 2000), while outside an action context human agents take a bottom-up approach driven by the features of objects and events available in their environment (Parkhurst et al., 2002). Thus, human attention can be modeled at two levels of abstraction: (1) at a low-level driven by object recognition; and (2) at a high-level driven by human intention.

Human intention refers to a human agent’s mental determination to act in a certain way. Human agents possess the ability to infer other human agent’s intentions by observing their actions (Blakemore and Decety, 2001). According to the simulation theory (Gallese and Goldman, 1998), a human agent A observing another human agent B’s action simulates the observed action as their own action and estimate the intentions for that action. Others
consider intention recognition to be an inferential process by applying a “theory of mind” (Brass et al., 2007). According to (Keysers and Gazzola, 2007), action simulation and inferential mechanisms are complementary in understanding action intentions. Irrespective of the actual mechanisms behind understanding action intentions which has created a lot of debate in the last few years, human intentions play an important influential role in human perception, action and allocating attention.

3.2 Basic Principles and Assumptions

3.2.1 Terminological Shift from “User” to “Agent”

The perspective in human–computer interaction research is consequently shifting towards conceiving human beings as mobile agents in a dynamically changing environment populated by physical and virtual objects alike (Pederson, 2003) instead of conceiving them as a “user” performing a dialogue with a computer. Within egocentric interaction, the term “human agent” is used instead of the term “user” for several reasons. Egocentric interaction models interaction occurring between a specific human being and potentially several interactive devices providing access to virtual objects at the same time, as well as the interaction between the human being and the physical environment populated with physical objects. In such a wide-ranging variety of interactive situations, it is more natural to regard the modeled individual as an agent in a physical-virtual environment (Pederson, 2003), i.e. environment populated with co-existing physical and virtual objects rather than a user of it. This shift in terminology, from “user” to “agent” is also a consequence of letting go of “the application” as the tacit focus: applications are “used” by someone and the “user” is defined by the artifact, whereas agents do actions and are engaged in activities that typically involve a number of objects. The notion of “agent” also implies that not all human actions need to be (explicitly or implicitly) directed towards a computing system. Egocentric interaction takes place within activity scenarios where only parts of the activity involve events that a computer system can or should care for. Compare the notions of “inband” and “out-of-band” in (Ullmer et al., 2005) for distinguishing between events that are part of a human-computer dialogue and those that are not.

3.2.2 Merging Virtuality, Mobility and Ubiquity Paradigms

The term “paradigm” is used in this thesis inspired by Thomas Kuhn’s influential notion of scientific paradigms (Kuhn, 1970). In analogy with scientific paradigms, we understand an interaction paradigm to include components such as: important design examples and use scenarios, important techniques and technologies, key problems and challenges, articulations of ideals and goals to pursue, interpretations of key concepts, such as ‘user,’
‘interface,’ ‘interaction,’ and finally, groups or communities of people (researchers and interaction designers) developing and defending the paradigm. Unlike Kuhn’s paradigms, different interaction paradigms may peacefully coexist; an older paradigm may find niches to survive the rise of a new dominating paradigm. Broadly speaking, there are three major interaction paradigms in operation at this time (or use paradigms as they are called in (Janlert, 2007)). In the older but still dominant virtuality paradigm, users interact with virtual objects, i.e., data objects, which are accessed through an interface that basically is a single stationary window to a virtual, symbolic world, usually with very little relation to the particular physical, real, use situation. In the newer ubiquity paradigm, human agents interact with multiple real objects augmented and interfaced by computer technology, which pervade the real world. In the also newer mobility paradigm, human agents move through the real world, while staying in contact and interacting with the virtual world through mobile devices. In one, now common scenario, human agents are using mobility primarily for remote access and operation, independent of the situation. In another, still atypical kind of scenario, mobility is used for in situ application, dependent of the situation; i.e., bringing the computational and informational resources to bear directly on the very situation of use in the real world. Egocentric interaction, it seems, is in effect merging the virtuality paradigm and its preoccupation with virtual objects with the mobility paradigm’s emphasis on movement and the human agent as a moving point-of-view in the physical world, and with the ubiquity paradigm’s acknowledgment of the importance of physical objects and the richness of the physical environment.

3.2.3 Different from Classical User-Centered Approaches

Clearly, the emerging new paradigm is very much centered on the “user,” or as we prefer to say, the (human) agent (3.1, above), not only in the sense that interaction design should bow to the user’s needs and preferences, should ultimately be a servant of the user, but in the sense that the agent should be understood as being in the (moving) center of the world. This distinguishes egocentric interaction from earlier “user-centered” approaches in HCI, such as, user-centered design (Norman and Draper, 1986), which largely ignored an agent’s current bodily situation (because it would have been fixed by the stationary computer system anyway), egocentric interaction acknowledges the primacy of the agent’s current bodily situation in the environment at each point in time in guiding and constraining the agent’s behavior (which is precisely what “egocentric” refers to). It assumes that the whole environment is taken into consideration, not just a single targeted artifact or system. It makes the assumption that proximity plays a fundamental role in determining what can be done, what events signify, and what the agent is up to. It seriously considers mobility and takes into account the agent’s more or less constant movements of head, arms, hands, and body, locally and through the environment, as well as the agent’s constant rearrangements and modifications of various parts of the environment. It is neither mainly
oriented towards the interaction with “virtual,” immaterial data objects (classical HCI), nor predominantly towards the interaction with physical objects and machines (classical ergonomics and Human-Machine Interaction), but pays equal attention to virtual and physical objects, circumstances, and agents, and their interrelations. It recognizes that the agent typically will have multiple ongoing activities at the same time, some of which may have little relation to each other, and that activities are started, put on hold, interrupted, resumed, and finished in a never ceasing flow. It makes no pretense and no tacit assumption that the system or the agent has full information of the situation, because situations are in principle open-ended and ever changing. It should be noted that there were and still are user-centered approaches to interaction design that do take some (but never quite all) of these factors into account, a prime example being participatory design (Ehn, 1989). However, these approaches are focused on methods and ideals for the design and development process, and on basically work-oriented and fixed, well-defined and stable tailor-made systems unlike computing systems for ambient intelligence. It should be noted that interaction paradigms that are centered on interaction devices will have to change frequently to keep pace with new technological advancements, while interaction paradigms that are human-centric like egocentric interaction are more stable over time: basic human characteristics change very slowly.

3.2.4 Situatedness: The View from Here

The following is a commonsensical summary of what we take to be the essence of the situatedness of agents with regard to interactions and activities, and which appears to be an important commonality for many of the theories and approaches referred to in section 2. The situative space model to be described in section 3.4 is based on the principle of situatedness. It is obvious that human beings have a self-centered perspective on their environment before any other perspective, such as second person perspective or some very-hard-to-acquire unlocated subspecie-aeterinitatis perspective that science aspires to (“the view from nowhere” (Nagel, 1986)). The world is so large and so rich in details that any agent with limited cognitive capacity must necessarily narrow its focus in some manner. Human beings have physical bodies that are located at a single particular place and oriented together with their limbs and sense organs in particular directions at any particular time. That gives them a natural primary vantage point for selecting which details and aspects to attend to: the view from here, what a particular embodied agent can perceive given its current bodily situation in the world. It also gives them a focus for actions: anything they can do is limited by what is within their reach given their current bodily situation in the world. To control your actions and the effects of your actions it is particularly helpful to be able to perceive the part of the environment in which you are acting and affecting. Vice versa, to perceive, to pick up information from a certain part of the environment it is helpful to be able to do actions that change your angle of perception and the parts you currently perceive because that can assist and improve the information extraction.
Natural agents are usually concerned about aspects and details of the world that have relevance for themselves, and naturally perform actions that are relevant for their own existence and relation to the environment. People are where they are, and that determines what they can do. If they want to do something that cannot be done right here right now, they will have to do something about it, and they will always have to do that starting from their very situation here and now. By doing things with what is within their reach, by moving themselves in the world, and by moving other objects in the world, they can change what they can do. The further removed from the current situation what they want to do is, the longer chains of action they need to be able to contemplate and follow to reach their goal. To their advantage, at each point in a long chain of actions, their current situation in the world, what they currently perceive, can help them recall and guide their next action; they may also be able to specifically arrange the environment around them and their relation to it so that their situational guidance through the process is improved. This being so, it also follows that the trajectory of a human body through space-time and the trail of environmental events and changes in its close vicinity may reveal a lot about which goals and intentions are on this human’s mind, and which activities are being carried out. Epistemic operations, i.e., physical operations performed to facilitate cognition rather than to further physical progress towards some external goal (Kirsh and Maglio, 1994), may offer particularly transparent access into the mind of the agent. Epistemic operations make cognition transparent by instantiating (partially) external cognition in the sense of the extended-mind hypothesis (Clark and Chalmers, 1998). They are transparent also in the practical sense of being within the scope of current and emerging tracking technology, just as other physical operations are. The view-from-here perspective means that all information about the environment that is picked up by the human agent derives from information converging on the body and modulated by the point of view of the body wherever it is situated; and that all actions, all effects on the environment produced by the human agent have their starting point in the body and are modulated by the angle of approach of the body wherever it is situated. Most human activities involve the handling of a number of objects and actions to make them available (locomotion, object transportation) in certain sequences and patterns that are appropriate for the objective of the activity. Human agents involved in a particular activity are at each moment guided in their actions by the array of objects in their immediate proximity, at the same time as some of their actions serve to or has as a side effect to change the array of objects in their immediate proximity. We may count on a general human tendency to minimize the effort spent, physical as well as cognitive. At any point in time, objects that are close are likely to be or become relevant for an ongoing or near future activity and objects that are relevant for an ongoing activity are likely to be or become close.
3.2.5 The Proximity Principle

Much of the enabling as well as limiting consequences of situatedness in this particular world we live in can be captured in the proximity principle (Janlert, 2006): Things that are close tend to matter; things that matter tend to become close. Objects (and other agents, circumstances) that are close to the human agent tend to matter in the sense that they have a fair chance of getting the human agent’s attention and figure in the human agent’s current cognitive processes and current activities, now and then also triggering (sometimes even forcing) action and the start of a new activity or the resumption of an activity currently on hold that directly involves the objects in case; and the closer the objects are the more likely they will play such a role, other things being equal. Objects (and other agents, circumstances) that matter to the human agent’s activities, first and foremost current and pending activities, will tend to either already be or soon become within close range; and the more imminent their use is, the closer they will tend to be. The rationale is obviously that if an object is needed in the activity the agent is likely to make sure that it is at hand in order to proceed with the activity, either by moving up to the object or by having moved the object to the current location in preparation (short term or long term). When a number of objects are needed simultaneously or in swift succession, we may consequently expect to observe prearranged environments for the purposed activities, as well as elaborated logistic strategies for moving objects around. The proximity principle aligns well with Satyanarayanan’s views on localized scalability (Satyanarayanan, 2001) which points out that “the intensity of communication between a human agent and the surrounding computing infrastructure will increase” within ambient ecologies and “if the communication and causality is not reduced based on locality metrics, we will get both system overload and user overload.”

3.2.6 Perception-Action instead of Input-Output

The classical HCI concepts of input and output need to be substituted with something that works both for physical and virtual object manipulation. Also, the concepts of input and output are device-centric and refer to device interfaces, while egocentric interaction needs to use concepts that refer to an agent’s interface to the environment (i.e., an agent’s input and output). (Streitz, 2007) discusses about the transition from human-computer interaction to human-environment interaction in the context of ambient intelligence. Hence, the concepts of perception and action replace the traditional concepts of input and output. Action is often inseparable from or intertwined with perception in physical everyday activities. One cannot change the state of an object (e.g., open a fridge door) without perceiving feedback (e.g., tactile feedback while opening the fridge door). Conversely, perception needs support of action (e.g., to see what is in the fridge, the door must be opened). This tight coupling between action and perception is the result of the way objects are designed, how we manipulate them, and how the manipulation process interplays with the laws of physics. One of the strengths of the direct manipulation mechanism (Shneiderman, 1983) widely
deployed in user interfaces for changing the state of virtual objects, is the fact that it makes the strong relationship between manipulation and perception prevail also in many parts of the virtual world.

The traditional concepts of input and output, stemming from a time when interaction typically was in terms of exchanges of language expressions, still have a bias towards a turn-taking approach where first an input is provided to a device that processes it and provides an output, whereas perception and action are inter-coupled and take place in parallel. One activity may typically involve several distinct sources (for perception) and targets (for action) at the same time, and multiple activities are often going on in parallel, more or less independent of each other.

### 3.2.7 The Physical-Virtual Equity Principle

Egocentric interaction differs from traditional interaction paradigms in explicitly ignoring the input and output devices of interactive computers such as keyboards and displays of PCs and cellular phones, considering them as more or less transparent mediators for accessing virtual objects. Mediators include sensors, actuators, input devices, output devices, enhanced by user-interface software, recognition algorithms, etc., that act as tools for providing human agents the access to virtual objects. Taking such a stance permits the modeling of physical and virtual objects as if they were situated in the same physical space, which is advantageous when modeling applications for ambient ecologies where the interaction complexity vastly surpasses what can be sufficiently described using a classical human-computer interaction dialogue model. We do recognize the inherent differences between physical and virtual objects (Pederson, 2003): the goal is not to make them resemble each other as much as possible. Designing virtual objects and their environments as an exact copy of the physical objects and their environments removes the inherent advantages of being virtual (or digital). For instance, it is easier and cheaper to transport a virtual object across space compared to transporting physical objects. Within egocentric interaction, we propose an approach where physical objects and virtual objects are co-located and complement each other with their inherent properties. Traditional interaction paradigms associate virtual objects to specific devices, while within egocentric interaction such an association is more dynamic and is often avoided in modeling human–environment interaction. The point with physical-virtual equity is to handle physical and virtual objects uniformly on a high level of abstraction in order to enable better modeling of mixed-reality situations. Furthermore, such a view aligns well with findings in psychology indicating that expert users of tools (whether a tennis racket or a computing device) tend to focus on the domain object they are working on (whether it is a tennis ball or an email) and become less aware of the details of the tool (such as the handle of the tennis racket or the workings of a particular input device) itself. Section 3.3 in this chapter is based on the physical-virtual equity principle.
3.2.8 Summary of Basic Principles and Assumptions

To summarize, the basic principles and assumptions of egocentric interaction are as follows: situatedness; attention to the local environment; the proximity principle; changeability of environment and agent–environment relationship; the physical-virtual equity principle; concept of perception and action; multiple concurrent activities; and incomplete information and control. From the assumption of situatedness, several principles have emerged like the proximity principle and human attention to the local environment. Human agents attend to their complete local environment even when interacting with a specific physical object; however they do so with varying levels of attention. Agents perform actions and engage in activities motivated and guided by what matters to them. The proximity principle has something to say about locality and how what matters is related to the relation between the agent and the environment, and so about what drives actions and movements. From that follows, among other things, that agents do manipulate objects and rearrange the environment in the pursuit of an ongoing activity, and that agents move to get close to objects (and places, and other agents) that are needed for an (upcoming phase) of some intended ongoing or upcoming activity. That accounts for some of the changeability of the environment and the relation of the agent to the environment; other agents are another source of change. The principle of physical-virtual equity is motivated pragmatically: our everyday interactions with virtual objects are already considerable and keep increasing. From the physical-virtual equity principle and the important fact that virtual objects are so much more mobile than physical objects (including the agent’s own body), we should expect an increase in the number of concurrent activities of an agent: while moving towards a physical object needed for one activity, it becomes possible to use the transportation time to make progress with other activities that in their current stage rely on virtual objects and actions; similarly, while waiting for the completion of an action directed towards some object involved in one activity, the waiting time can be used for virtual handling of other activities. The incompleteness of information and control, finally, implying an element of improvisation that has been absent or suppressed in earlier paradigms, follows from the changeability of the environment and the agent–environment relationship, and from the richness of the real world.

3.3 A Physical-Virtual Design Perspective

As mentioned earlier, this thesis envisions ambient intelligence along mixed-reality environments where physical objects and virtual objects coexist. An attempt to integrate the physical and the virtual is addressed by Thomas Pederson in his PhD thesis (Pederson, 2003) and this section attempts to build upon the concepts presented in his work. The physical-virtual equity principle described in section 3.3 motivates the exploration of a physical-virtual design perspective for ambient intelligence. The focus is on the conceptual design perspective that enables building ambient ecologies intended to relieve the occupants from some of the extra efforts currently
needed when performing activities that make heavy use of both physical and virtual aspects. There are at least three motives for adopting such a physical-virtual design perspective: (1) Human agents well acquainted with specific (physical and virtual) environments are typically more concerned with the manipulation of (physical and virtual) objects than the user interface through which they are accessed. (2) Such a design stance facilitates the conceptualization of objects that bridge the gap between the physical and virtual aspects of an ambient ecology. (3) Many existing (physical and virtual) objects already have both physical and virtual manifestations (e.g. a printed photograph within a photo frame and its digital version within a laptop) thereby introducing the possibility of seamlessly crossing the physical-virtual boundary. The physical-virtual approach is intended not only to integrate physical and virtual objects, but also to explore the possibilities of integrating physical and virtual situations in which occupants of an ambient ecology might find themselves in the future. Many existing researches view ambient ecologies as physical environments augmented with computing technologies. Such a bias towards the physical aspects of an ambient ecology introduces additional cost for its occupants to access its virtual aspects, and to move across the physical-virtual boundary, effectively creating a physical-virtual environment gap (Pederson, 2003).

3.3.1 Physical-Virtual Environments

Physical environments generally possess some basic relations and ordering among self-sustained physical objects, and are governed by the laws of physics. They usually possess rich representations and offer natural affordances. In contrast, virtual environments are dependent on the mediators for their existence and do not have universal laws. Virtual environments introduce inexpensive space navigation and object transportation, allow for making big changes and even reversing operations with ease compared to the physical environments (Pederson, 2003). Virtual environments offer the possibilities of multiple manifestations of virtual objects in distinction to physical environments that only offer one unique manifestation of every single physical object. The fact that physical environments and virtual environments are inherently different introduces the possibility of reaping the best of both environments by allowing them to co-exist and complement each other. However, physical environments and virtual environments are similar in certain respects: they both allow a human agent to be situated in them and interact with objects contained in them, thus providing an arena for their activities. This makes it possible to view them as being alike and merge them together as physical-virtual environments where physical objects and virtual objects co-exist. In this thesis, the term physical-virtual environments or mixed-reality environments (Costanza et al., 2009) is used to represent environments that are part of the real world and in parallel include virtual spaces wherein human activities takes place both in the physical world and in the virtual world (Pederson, 2003). In this section, the physical-virtual design perspective will be discussed with examples from an ambient ecology within a home context.
According to (Pederson, 2003), traditional interaction paradigms do not address the challenge of integrating the physical world and the virtual world (world created using computing technology), while within a physical-virtual environment there arises a need to integrate the physical world with the virtual world from a human agent’s perspective. The egocentric interaction paradigm proposed in this chapter attempts to reduce the gap between the physical and the virtual by taking a physical-virtual design perspective. According to (Buxton, 2002), “One of the most significant issues confronting computer users, is the problem of bridging the gap between the physical and virtual worlds. For most activities, most current systems make it too difficult to move the artefacts back and forth between these two worlds, the physical and the virtual. Hence, the relevant documents, designs, etc. are isolated in one or the other, or split between the two.”

3.3.1.1 Presence within Multiple Environments

Human agents could subjectively be present in more than one environment at a time, in parallel. Instead of forcing human agents to shift between different environments unnaturally, ambient ecologies attempt to merge the physical and the virtual environments into one global experience where the human agent can shift between the environments in two different ways.

In the first approach, human agents are present in one environment at a time. Their attention switches along with the sequence of environment shifts. The switching should be “natural” and “intuitive.” In the second approach, human agents are spatially present in several environments at a time with varying levels of attention and intention within individual environments. Depending on the human agent’s context, the ambient ecology could decide on the appropriate approach to take.

In both approaches, the merging of the physical and the virtual environments should be dynamic in order to retain the benefits of the rapid changes that the virtual environments allow. Even though in theory there exists a difference between the above mentioned two approaches, in practice a human agent might only be able to really focus on one thing at a time, thus ending up in the first approach to some extent. The idea with the second approach is that a human agent might be able to fully focus on only one task (or environment) at a time, but might still be able to partially perceive information through their peripheral attention. This is clearly an interesting issue, both conceptually and implementation-wise and needs further research.

3.3.1.2 Mobility

The effect of a human agent’s mobility within physical-virtual environments introduces several challenges. Consider a scenario where a human agent is moving while in a physical environment performing actions in a virtual environment without any mobile/wearable mediator. It introduces a challenge in presenting the required virtual environments in a stable manner. The
virtual environments are dependent on the mediators that are available in the physical environment in which the human agent is located. Some environments might have many mediators that are conducive for presenting the virtual environments, while some environments might be dumb with no mediator like in the middle of a forest. This issue is mentioned as uneven conditioning in (Satyanarayanan, 2001). Since virtual objects are flexible and manifest themselves in different modalities (visual, audial, tactile), virtual environments could be presented in both mediator rich and mediator deficient environments. Mediators that are always on and worn by a human agent offer an alternative solution in maintaining a stable virtual environment. A human agent's navigation within virtual environments often has little consequence in the physical environment. The physical environment remains more or less stable.

The intensity of a human agent's interaction with physical objects while moving is usually low. For instance, human agents navigate to different rooms, but perform more intensive actions within individual rooms compared to the actions performed (in terms of interaction with objects) while navigating between them. This might be due to the fact that a major proportion of human perception and action abilities are spent on navigation. Taking inspiration from the human agents’ mobility within physical environments where their interaction with objects is usually of low intensity, it is proposed to facilitate low-intensity interaction with virtual objects mainly through wearable mediators while the human agents are moving in the physical environment. Such an approach might reduce the number of virtual objects that could be presented (presented mainly through wearable mediators) but introduces better stability in presenting virtual objects. Mediators that quickly enter and leave a human agent’s physical environment are simply ignored for simplicity.

### 3.3.1.3 Entering and Leaving Environments

Human agent's mobility in the physical world is often used as a common strategy to enter or to leave physical environments. For instance, human agents leave their home, enter a supermarket, leave the supermarket and enter their home again. In principle, human agents are situated in one physical environment at a time and as they leave one physical environment, they enter another physical environment. Since such environments afford a set of physical activities each, by entering or leaving an environment, a human agent invariably switches between activities as well. Of course, there are activities that are less dependent on specific environments like running that are ignored in this discussion.

There are some activities that span across more than one environment and often require shifts between different environments. As mentioned earlier, a human agent moves lightly in the virtual world and it is often possible to enter or leave more than one virtual environments at the same time. Traditional approaches to entering and leaving virtual environments are by starting an application that brings up a window on the desktop and to
close the application, both of which are controlled by a human agent. Ambient ecologies might include many applications that are closely integrated to the physical environment and needs to be proactively started and closed providing a feel of human agents entering and leaving specific physical-virtual environments. Further research on strategies to facilitate the entering and the leaving of physical-virtual environments (or mixed-reality environments) is required. One approach is to take inspiration from physical entrance and exit, and make it similar for physical-virtual environments where the physical environment to some extend determine the virtual environments that could potentially be integrated in creating a physical-virtual environment.

3.3.1.4 Centricity

Physical-virtual environments introduce the centricity problem, i.e. to choose an appropriate viewpoint for the observer relative to the objects of interest (Milgram and Kishino, 1994). Physical environments are usually viewed from a human agent’s egocentric perspective based on their situatedness and embodiment, and this makes it easier for a designer to deal with physical environments. However, virtual environments could be presented from multiple viewpoints at the same time for individual agents since virtual environments are artificially created using computing technology. This makes it harder for a designer to decide on the viewpoint and to handle virtual environments in terms of their presentation, affordance, etc. even though a designer has more choices in terms of dealing with virtual environments. Integrating the physical and the virtual environments makes it even more challenging for a designer in establishing appropriate centricity within physical-virtual environments.

3.3.1.5 Access and Control

The rules for having access to and controlling resources in a physical environment is more well established compared to similar rules within virtual environments. For instance, proximity is often used as an implicit rule to access the resources in a physical environment. The person closer to a chair has a better opportunity to use it compared to someone at a distance. Similarly, ownership could be considered as another implicit rule where a person who owns an apartment has more access and control rights compared to a guest in that apartment. There are rules in virtual environments as well, like guest rights, owner rights, moderator rights, etc. But human agents within physical-virtual environments might be in the same physical environment, yet access and control different virtual environments. For instance, human agent A might read the morning news (a virtual object) through the dining table (a mediator), while human agent B might check her emails (also a virtual object) through the dining table. Similar to the centricity problem within physical-virtual environments, the access and control rules are a major challenge that needs further investigation.
3.3.1.6 Privacy

Physical environments protect the human agents’ privacy physically, by having doors for instance, so that they can perform both private and non-private activities. Human agents themselves are usually responsible for privacy across the social dimension. Virtual environments usually have access modifiers like private, public, protected, etc., to secure users’ privacy. Within physical-virtual environments, protecting the human agents’ privacy can be a challenge especially since the physical and the virtual aspects are integrated requiring further research.

3.3.2 Physical-Virtual Artefacts

Physical objects are manifested within a physical environment, while virtual objects are manifested within the realm of a computer screen; this is the traditional view. In a physical-virtual environment, however, both physical and virtual objects possess physical-virtual manifestation possibilities, i.e. they can be manifested both physically and virtually. Such objects are referred to as physical-virtual artefacts (PVA) in (Pederson, 2003): “a physical-virtual artefact is an abstract artefact that is manifested in both the physical and the virtual environment, where these manifestations to a large extent utilize the unique affordances and constraints that the two different environments facilitate, and finally where one manifestation of a specific physical-virtual artefact is easily identified if a corresponding manifestation in the other environment is known.” Refer to Fig. 3.1a.

- Example 1: A physical stove having a virtual manifestation that can be accessed from the living room through the wall display and a gesture recognizer that acts as mediators. The virtual manifestation allows the human agents to perceive the state of the stove and manipulate it from a distance.
- Example 2: A virtual media player that can be accessed through a dining table and a candle holder that acts as mediators. By manipulating the physical candle holder (when not in use), different music play lists can be selected and played. Rotating the candle holder will increase or decrease the sound volume.

Physical-virtual manifestation of objects allows the human agents to interact with such objects (irrespective of being physical or virtual) both physically and virtually depending upon the usage context thereby allowing them to be flexible. In some situations, the physical affordance of an object might be preferred over their virtual representation, while in other situations the light-weight of being a virtual representation might be preferred over their physical representation by human agents. For instance, a human agent located closer to a stove in the kitchen might prefer to turn the stove off using a physical knob considering the handy affordance that it offers; while the same human agent when in the living room might prefer using a virtual knob on the wall display to avoid locomotion. Physical-virtual manifestation of
objects also allows for integrating the physical and the virtual aspects of an
environment since the state of an object irrespective of its manifestation
(physical or virtual) is synchronized, and physical-virtual artefacts enable
simpler shifting between the physical and the virtual aspects of an
environment.

3.3.2.1 Direct and Mediated Access

Traditionally, objects belonging to physical environments are accessed
directly by human agents without the need for mediators (i.e. objects that
mediate events in the physical and the virtual environments), which are
referred to as direct access. There are some objects in the physical
environment that might be distributed like a ceiling lamp and a switch which
are physically distributed but could be combined and considered as a single
artifact. Such objects are referred to as distributed artefacts in (Pederson,
2003) and do not require mediators for a human agent to interact with them.
With the augmentation of computing technologies within everyday objects,
clothes and environments, physical objects could be accessed virtually
through mediators from a distance referred to as mediated access.

Virtual objects are artificially simulated computing objects that do not
exist in physical reality necessitating the use of mediators to access them in
physical environments.

- **Example of direct access:** A human agent physically unlocks and opens
the entrance door of their home for the evening’s guest.

- **Example of mediated access to physical object:** A human agent while
being busy in the kitchen unlocks and opens the entrance door of their
home for the evening’s guest through the touch screen on the
thermometer display.

- **Example of mediated access to virtual object:** A human agent uses
their tooth brush in the morning in front of a bathroom mirror
initiating the personalized news client virtual object to be initiated
and present the latest news on the bathroom mirror to the human
agent.

- **Another example of mediated access to virtual object:** A human agent
uses the coffee cup on their table to adjust the volume of the music
jack virtual object.

From a human agent’s perspective, mediators should be subjectively
transparent in order to facilitate the co-existence of physical and virtual
objects within physical-virtual environments. Transparency could be physical
where the mediators are so small that they are physically invisible, could be
cognitive where the mediators do not demand any additional attention
requirements while mediating interaction between a human agent and the
domain objects, could be virtual where the mediators are implicit without
demanding explicit human interaction, etc. It is less important to distinguish
between physical and virtual objects, but more important that human agents
are aware of and exploit the differences in their characteristics. It is
important that objects (physical or virtual) in an environment provide the appropriate affordance to perform activities. Hence by introducing the concept of mediators that are transparent (also refer to the physical-virtual equity principle), the modeling of a physical-virtual environment that enables human agents to interact with domain objects across the physical-virtual boundary is simplified: it only needs to care for the domain objects.

### 3.3.2.2 Relationship between Mediators and Domain Objects

The relationship between mediators and domain objects (physical or virtual) can be both one-to-many and many-to-one. For instance, a refrigerator display (a mediator) might provide access to many virtual domain objects like the temperature controller, old food detector and the shopping assistant. Similarly, the shopping assistant (a domain object) might be accessed through many mediators including the human agent’s wristwatch, the refrigerator display, and the entrance door display and aurally through the radio.

### 3.3.2.3 Perception and Action

Human agents of a traditional physical home environment interact with physical objects through processes of perception and action. When it comes to interacting with virtual objects, however, the classical HCI concept of input and output is normally used which is well established in the WIMP interaction paradigm. The input and output concept becomes inadequate within physical-virtual environments, where the human agents are expected to interact with multiple devices, often in parallel. Other issues include the human agents’ attention and intention towards interaction with individual devices. The human agents’ attention and intention may vary continuously depending upon the context of interaction with individual devices.

Within the proposed physical-virtual design perspective, a human agent’s interaction with virtual objects is also considered to be based on their perception and action processes. Such an approach allows for an egalitarian stance in facilitating interacting with domain objects, irrespective of being physical or virtual. It also allows the human agents to interact with multiple domain objects (both physical and virtual) at the same time with different levels of attention towards the individual domain objects.

In section 3.5, two spaces namely the perception space and the action space are introduced as part of a situative space model. The perception space contains a set of domain objects that are perceivable by an individual human agent at a particular moment in time while the action space contains a set of domain objects that are manipulable (or actable) at a particular moment in time. Due to the fact that human agents are mobile within physical-virtual environments, the domain objects within the perception space and the action space keep changing dynamically. Also, the processes of perception and action take place in multiple modalities including visual, aural, and tactile.
modalities thereby requiring a multimodal approach to defining the perception and the action spaces.

### 3.3.3 Physical-Virtual Situations

Awareness about a human agent's situation is important for ambient intelligence to be able to adapt and respond to the human agent's immediate needs and to facilitate their current activities (Mastrogiovanni et al., 2010). Traditional approaches consider a human agent's physical situation in adapting either the virtual environment (as in context-aware computing) or the physical environment (for instance, as in home automation).

- **Example of virtual environment adaptation:** A virtual cooking guide detects what a human agent is currently cooking and provides suggestions according to the physical activity context.
- **Example of physical environment adaptation:** As a human agent wakes up in the morning the coffee maker in the kitchen begins to automatically prepare coffee.

A human agent’s virtual situation is often ignored in adapting the human agent’s physical environment or their virtual environment. For instance, one could imagine that depending on whether a human agent is sending an email or is watching an online movie, the room lighting could adapt and provide an ambient experience. Within the physical-virtual design perspective, a human agent’s situation that spans across the physical-virtual boundary is given importance while being aware of the inherent differences between physical and virtual situations. For instance, a human agent’s proximity to a set of physical objects might be more stable compared to a human agent’s proximity to a set of virtual objects that might appear and disappear as a result of dynamic switching between different virtual environments. The intention is to handle physical and virtual situations uniformly at a higher level of abstraction in order to better model a human agent's physical-virtual situation.

Situation models that acquire knowledge through observation and evolve during use are important for ambient ecologies. Existing situation models predominantly focus on either the physical aspects (Mastrogiovanni et al., 2010) or the virtual aspects of an ambient ecology. The Situative Space Model to be described in section 3.5 is physical-virtual, i.e. covers both the physical and the virtual aspects of a human agent’s situation.

### 3.3.4 Physical-Virtual Activities

Awareness about a human agent’s current activities is important for providing support to those activities. However, a human agent’s activity may span across the physical-virtual boundary introducing a need to recognize activities both in the physical environment and in the virtual environments. Recognizing a human agent’s virtual activities is simpler in comparison to
recognizing their physical activities. Virtual activities takes place within virtual or computer simulated environments with clearer events and are not affected by the uncertainties that exists in the real (physical) world. Also, physical activities are detected using sensors that might be unreliable with their inherent limitations.

Human activities could be represented using three levels of abstraction: activity, action and operation, inspired by activity theory (Kuutti, 1996). An activity has an objective and is comprised by a set of actions that have well-defined goals and are accomplished by largely unconscious operations. Both actions and operations can be either physical or virtual. Activities that contain only physical actions are referred to as physical activities, while activities that contain only virtual actions are referred to as virtual activities. A third type of activity that is common in physical-virtual environments and contains both physical actions and virtual actions are referred to as “physical-virtual” activities (Pederson, 2003). Physical-virtual activities introduce frequent switching between the physical and the virtual aspects of a human agent’s environment and will be one of the exploration focuses within the context of this thesis. One reason why human agents would perform physical-virtual activities is because by switching between physical and virtual environments, human agents will be able to get (predicted by the human agent) better support for their next action.

3.4 Situative Space Model

The situative space model is motivated considering the principle of situatedness that shreds light on the human agent’s situation. The situative space model is intended to capture what a specific human agent can perceive and not perceive, reach and not reach, at any given moment in time (Fig. 3.1). It has been inspired by cognitive-science theories relating to context and situatedness (Nardi, 1996b). This model is for the emerging egocentric interaction what the virtual desktop is for the WIMP interaction paradigm: more or less everything of interest to a specific human agent is assumed to, and supposed to, happen here. Although spatial and topological relationships between objects within a particular space certainly are of interest, within the context of this thesis, this research so far has mainly taken into account whether an object is present in a space or set, or not. The state and state changes to objects are important and are taken care of in recognizing human activities and in facilitating a human agent’s interaction within ambient ecologies, however object states and their changes are beyond the scope of the situative space model which is intended to be simple. Applying the model in this simple way generates a number of objects for each space and set, at any given time instant. The situative space model is inspired by the proximity principle (refer to section 3.2.5) which makes the assumption that proximity plays a fundamental role in determining what can be done, what events signify, and what human agents are up to. The situative space model considers the physical-virtual equity principle in framing both physical and virtual objects within the individual spaces and sets.
In Fig. 3.1, the spaces represent presence and approximate spatial relationship among physical and virtual objects with respect to what a specific human agent can perceive (perception space) and manipulate (action space) at a given moment in time. Whether objects are perceivable and manipulable depend on their relations to the human agent in all available interaction modalities, e.g. vision, touch, audio (Pederson et al., 2011).

### 3.4.1 Main Components of the Model

The following definitions are agent-centered but not subjective; they are principally aimed at allowing objective determination and thus are suitable for automated tracking purposes. Refer to chapter 5 and 6 for further information about the tracking of the situative spaces. Chapter 7 describes how this model is used for activity recognition while chapter 8 describes how this model is used for facilitating egocentric interaction within the easy ADL ecology.

#### 3.4.1.1 World Space (WS)

The world space refers to the space containing the set of all physical and virtual objects to be part of a specific model. From an operational perspective, it refers to the physical and virtual objects that could be accessed within an ambient ecology.
3.4.1.2 Perception Space (PS)

The perception space refers to the part of the space around the agent that can be perceived at each moment. Like all the spaces and sets defined below, it is agent-centered, varying continuously with the agent's movements of body and body parts. Perception Space can be given a simple geometrical interpretation (like a cone, in the case of vision, e.g.) as a rough approximation. Objects may occlude other objects and thus create (temporary) holes in the space. Different senses have differently shaped perception spaces, with different operating requirements, range, and spatial and directional resolution with regard to the perceived sources of the sense data. Compare vision and hearing, e.g.: the perception space of vision requires light, is basically cone-shaped, with in principle infinite depth range if there are no obstructing objects, very good angular resolution and fairly good depth resolution at close range; the perception space of hearing requires air (or some similar medium), is basically ball-shaped, with quite limited range, good angular resolution for higher pitches, low resolution for low pitches, and rather poor depth resolution. You cannot see what is behind your back, but you might hear it; on the other hand, many objects are silent (and, contrary to how vision works, offer little object-specific information by way of modulating sound from other sources at the scene) but can be seen. The different perception spaces of different senses complement each other.

The Perception Space in our definition can either be interpreted as the complex superposition of the perception spaces of the different senses, or it can be interpreted as dealing with each perception space in isolation. At this time we want to specifically focus on Visual Perception Space, and some of the definitions below may need revision when considering other perception spaces. Within Perception Space, an object may be too far away to be possible to recognize and identify. As the agent and the object come closer to each other (either by object movement, agent movement, or both) at some point, at some distance, the agent will be able to identify it as X, where X is a certain type of object, or possibly a unique individual.

A particular object can be of several different types, e.g., with different levels of abstraction (my car, a Toyota, a car, a moving object, etc.), but for a particular type X the distance at which it can be perceived as X can approximately be related to attributes of X such as size and presence of distinguishing perceptible features. For vision, also viewing angle may be important; many objects are difficult to recognize from certain angles. In the dynamics of a certain situation the agent will be able to compensate by changing the viewing angle (by head movements, by locomotion, by waiting for the object to turn, by actively turning the object). Increasing the distance again, or changing the viewing angle, the perception of the object, although not sufficient to recognize the object as of type X, may still be able to serve as a token, a perceptual reminder of its type. For each type X, the predicate “perceptible-as-X” will cut out a sector of Perception Space, the distance to the farthest part of which will be called recognition distance.
3.4.1.3 Recognizable Set (RS)

The recognizable set refers to the set of objects currently within Perception Space that are within their recognition distances. The kind of object types we are particularly interested in here are object types that can be directly associated with activities of the agent – ongoing activities, and activities potentially interesting to start up – which is related to what in folk-taxonomy studies is known as the basic level (Rosch, 1978). This is the level of a hierarchical taxonomy at which within-category similarities are maximized and between-category similarities are minimized. Objects belonging to one and the same category at the basic level are typically similar and distinctive in appearance and can also be associated with similar and distinctive motor operations. The basic level represents in a sense the basic operative level of human activities. E.g., when we are thinking of activities involving hand tools, relevant basic level object types would be hammer, saw, screw driver, etc.; each type easily recognized from its distinctive visual appearance.

To perceive the status of a designed object with regard to its operationally relevant (perceivable) states (operations and functions as defined by the designer of the artifact) it will often have to be closer to the agent than its recognition distance: the outer limit will be called examination distance.

3.4.1.4 Examinable Set (ES)

The examinable set refers to the set of objects currently within Perception Space that are within their examination distances. Normally, we expect the Examinable Set to be a proper subset of the Recognizable Set.

3.4.1.5 Action Space (AS)

The action space refers to the part of the space around the agent that is currently accessible to the agent’s physical actions. Objects within this space can be directly acted on. The outer range limit is less dependent on object type than PS, RS and ES, and is basically determined by the physical reach of the agent, but obviously depends qualitatively also on the type of action and the physical properties of objects involved; e.g., an object may be too heavy to handle with outstretched arms. Since many actions require perception to be efficient or even effective at all, Action Space is qualitatively affected also by the current shape of Perception Space.

From the point of view of what at this stage can be relatively easily automatically tracked on a finer time scale, it will be useful to introduce a couple of narrowly focused and highly dynamic sets within Action Space (real and mediated).
3.4.1.6 Selected Set (SdS)

The selected set refers to the set of objects currently being physically or virtually handled (touched, gripped; or selected in the virtual sense) by the agent.

3.4.1.7 Manipulated Set (MdS)

The manipulated set refers to the set of objects whose states (external as well as internal) are currently in the process of being changed by the agent. Normally, we expect the Manipulated Set to be a subset of the Selected Set.

For all of the above defined spaces and sets, geometrically defined sectors and object-type-dependent memberships are in principal computable, which, together with current state-of-the-art sensor technology, makes it possible to automatically track their contents without requiring an elaborate real-time model of the agent’s cognitive processes. Clearly, that an object is known to be within the visual Perception Space, e.g., is still no guarantee that it actually has been perceived or that it will be. All these spaces and sets, with the obvious exception of the Selected Set and the Manipulated Set, primarily provide data on what is potentially involved in the agent’s current activities. They are still quite useful in creating a first rough approximation of what is going on – good enough to make usable detections and predictions of ongoing and upcoming actions and activities, as we will see in the prototyping experiments reported in chapter 7. The proposed situative space model is used for activity recognition (refer to chapter 7) and for interaction management (refer to chapter 8).

3.5 Applying the Situative Space Model

The following analysis is presented partly to make the theoretical description of the SSM in the previous section more concrete, partly to indicate how the model could be used in the design phase for empirical analysis of mixed-reality ambient ecologies.

In Fig. 3.2, table (P18) has visual display (M1) currently showing information from a diet application (V1). The wall calendar (P28) has a visual display (M10) and a touch sensitive surface (M11) currently showing and providing access to a calendar application (V2). The wireless headset (P31) in the human agent’s right contains a microphone, earphone, button, and LED (M2-M5). The cellular phone (P30) in the trouser pocket includes a visual display, keyboard, earphone, loudspeaker, and microphone (M6-M9, M14). All mediators offer means for explicit interaction between the human agent and the virtual objects. Also highlighted in the figure is a glass of milk (P1) currently being manipulated by the human agent, as well as a piece of bread (P2) in front of him. (Other physical objects lack labels in order to keep the figure simple.)
Fig. 3.2. A human agent having breakfast situation (Pederson et al., 2011).

3.5.1 Situation

A human agent sits down at the kitchen table in order to have breakfast. The kitchen table is fitted with a visual display in the centre of the tabletop. In his pocket he has a cellular phone and on his right ear a wireless headset. A wall calendar two meters away has an embedded touch screen. Various software applications are running on a server ready to interact with the human agent through these mediators. Fig. 3.4 illustrates this scene with the mediators.
and a few objects highlighted. Fig. 3.3 shows the situative space model applied to the same situation.

**Fig. 3.3.** The breakfast situation of Fig. 3.2 as viewed through the situative space model (Pederson et al., 2011).

In Fig. 3.3, some virtual objects (V3-V13) not visible in Fig. 3.2 are shown here in the world space, ready to be made accessible to the human agent through mediators in the perception and action spaces. Flows of interaction – specifically, manipulation of virtual objects and perception of the results – are illustrated by arrows. Lines without arrowheads indicate more static relationships among objects (Pederson et al., 2011).
3.6 Associating the SSM to Human Intention and Attention

![Geometrical representation of the situative space model, and its association to human attention and intention](image)

**Fig. 3.4.** A geometrical representation of the situative space model, and its association to human attention and intention (in terms of activity and action context) represented in set theory notations (Surie et al., 2012).

Associating the situative space model to human intention and attention (refer to Fig. 3.4) is useful in facilitating a human agent’s interaction within ambient ecologies. Keeping track of all the objects (physical objects and computer-generated virtual objects) within a human agent’s perception and action space can be cognitively expensive for a human agent. It is assumed that a human agent scans the perception and action space to find objects that are relevant for their intentions. Refer to section 3.1.4. The term intention might have several interpretations, but we focus on human intentions that are possible to infer from recognition of the accompanying activities. An important feature of the egocentric interaction paradigm is to unobtrusively sense and infer human intentions (activity and action context) and adapt a human agent’s interaction experience with post-desktop computing environments. Human activities and actions can be modeled and recognized...
by keeping track of the changes in the content of the situative space over time (refer to chapter 8), while their operations could be tracked by capturing the events that occur within the selected set and the manipulated set (refer to chapter 6) which are part of a human agent’s action space.

Human agents usually perform activities and actions by interacting with and manipulating objects over time. Since the situative spaces are described in terms of physical and virtual objects, relationships between individual objects and their association to individual human activities and actions can be established. Such an association can be achieved in at least two ways. One approach is to empirically collect data by observing human activities and their relationship to objects. Activity recognition systems based on a human agent’s interaction with objects usually possess techniques for automatically collecting empirical data about an object’s association to individual activities. Another approach is more a top-down approach based on the artifact designers’ assumption of the activities for which individual objects are designed.

By modeling and recognizing human activities and actions, one could distinguish intentional operations from unintentional operations (which could be regarded as noise) that are common in real-world scenarios. Also the higher-level knowledge about the human agent’s current intentions enable filtering out activity irrelevant objects from being present within the human agent’s perception and action spaces, to reduce the human agent’s attention and cognitive requirements. Such higher-level knowledge could also be considered as implicit input (Schmidt, 2000) to computing applications thereby removing the need to bother a human agent by requesting for explicit inputs.

Human attention is an important and scarce resource. Ambient ecologies might include many computers trying to interact with a human agent in parallel as with Mark Weiser’s vision of ubiquitous computing (Weiser, 1995). In such situations it is important not to overload a human agent’s attentive and cognitive capabilities. Human agents use different levels of attention within the perception and action spaces. The different levels of attention include no attention, peripheral attention, central attention and agent response. The benefit of considering a human agent’s peripheral attention is due to the fact that human agents could perceive a lot of information through peripheral attention without disturbing their limited central attention capabilities. Such a pragmatic view of a human agent’s attention provides the possibilities of seamlessly moving virtual objects between the center and periphery of human attention (Weiser and Brown, 1997).

In ambient ecologies, virtual objects could be positioned in a human agent’s situative spaces using a mixed-initiative approach (Horvitz, 1999). Human agents could directly manipulate virtual objects and position them within the situative spaces and to complement it, virtual objects could be automatically positioned within the situative spaces through artificial software agents like the interaction manager to be described in the next section. Physical objects usually exist in a physical environment and are usually not required to be artificially positioned in the situative space with
the exception of having artificial physical agents like robots that might bring and position physical objects. Instead, the physical objects are usually sensed to determine their position in the situative spaces which are used for managing a human agent’s interaction with computing applications that make use of the physical objects and their context. Human attention to an extent is dependent on the objects in the situative spaces. Some object might grab human attention, while other might reside peacefully in the periphery of human attention. An attempt to pragmatically associate human attention to the situative spaces is as follows (bottom-up approach). The association is more of a rule of thumb for cognitive systems.

In Fig. 3.7, the various levels of human attention are mapped to different sets and spaces within the situative space model.

- **Objects in the world space**, but outside the perception and action space demand no attention.
- **Objects in the perception space**, but outside the recognizable and examinable set demand peripheral attention.
- Objects in the recognizable and examinable set demand central attention.
- **Objects in the action space**, but outside the selected and manipulated set demand agent response, i.e. to manipulate them (physically and/or virtually).
- **Objects in the selected and manipulated set** are already being acted upon and is assumed to demand central attention.

Human attention is also dependent on human intentions. An object that is currently associated with the human agent’s current activity has a better chance of occupying a human agent’s peripheral attention than objects that are outside the activity context. Similarly, objects associated with the human agent’s current action have a better chance of occupying a human agent’s central attention. Refer to Fig. 3.4. An attempt to pragmatically associate human intention to the situative spaces is as follows (top-down approach):

- **Objects outside a human agent’s activity and action context** demand no attention and the object is pushed outside the perception space; and no agent response is required and the object is pushed outside the action space.
- **Objects within a human agent’s activity context, but outside the action context** demand peripheral attention and the object is pushed inside the perception space, but outside the recognizable and examinable set; and agent response is required and the object is positioned in the action space but outside the selected and the manipulated set.
- **Objects within the activity and the action context** demand central attention and the object is pushed inside the recognizable and/or the examinable sets; and agent response is required and the object is positioned in the action space but outside the selected and the manipulated set.
3.7 Discussion

The proposed egocentric interaction is centered on a human agent and in particular their cognition. This introduces two types of challenges: 1) to build theoretical models of a human agent and their cognition; and 2) to develop technologies that can accurately sense, recognize and model aspects of a human agent and their cognition.

Egocentric interaction is intended for everyday environments that could be occupied by multiple human agents. The proposed situative space model should be extended along the following dimensions: a) to consider social and cultural aspects of a human agent; b) to be a part of a network of situative spaces formed by multiple occupants of an everyday environment. Such a network might include situative spaces that are shared among multiple human agents useful in facilitating collaborative activities; and c) to include spatial relationship among objects within the situative spaces as it plays an important role in human behavior. The assumptions upon which the egocentric interaction paradigm is developed might not be valid in some situations, even though they are useful for most situations. For instance, the proximity principle that is used in operationalizing the situative space model might conflict in situations where a far-away object might be more important than the ones that are closer to the human agent. Such exceptions should be explored and handled in the future.

Egocentric interaction is dependent on technical and technological advancements. Accurate sensing of the situative spaces in everyday environments subject to noise, sensor failures and dynamic changes to the situative spaces is an important challenge to address. Modeling and recognizing activities in the real world is another important challenge to address. Human activities are usually performed with variations; they change over time, and are often interleaved with other parallel activities. Human activities might take place in a mobile context, and get constantly interrupted and resumed without noticeable events that signify it. Also, multiple human agents might perform group activities, or attempt to share resources in performing their individual activities resulting in conflicts. Accurate and fine-grained activity recognition at the action level is important for facilitating egocentric interaction. Presenting and providing access to virtual objects within the situative spaces is dependent on the available mediators within the respective situative spaces. Different environments might have varied riches in the quantity and quality of available mediators and the interaction manager should handle even the worst situations where limited mediators are available for enabling a human agent’s interaction with computing applications. Refer to chapters 4 to 8 for further information on the technical and the technological developments for facilitating egocentric interaction within ambient ecologies. It is acknowledged that further work is required not only from a technological perspective, but also in designing experiments so as to externally validate (Mitchell and Jolley, 2001) the sensor technologies, the recognition algorithms, and the concepts that form a theoretical base for the emerging egocentric interaction paradigm.
To conclude, egocentric interaction has its foundations on a human agent and their cognition, and is expected to withstand (and benefit out of) the technological advancements in evolving as a suitable interaction paradigm for ambient intelligence.
Chapter 4

The easy ADL Ecology: an Infrastructure for Ambient Intelligence

This chapter will describe the easy ADL ecology, an infrastructure for ambient intelligence built based on egocentric interaction with specific focus on the situative space model and a physical-virtual design perspective described in chapter 3. The easy ADL ecology consists of smart objects, a personal activity-centric middleware, ambient intelligence applications and a human agent in the middle of it all. The easy ADL ecology was first implemented in an immersive virtual reality home environment and then in a living laboratory home environment. The easy ADL ecology supports a single human agent with their everyday activity support. Interaction management rules and techniques to facilitate human interaction with virtual objects in the easy ADL ecology are also described in this chapter.

4.1 Introduction

As mentioned earlier, with the advancement in computing, communication, sensing, actuation, interface and interaction technologies, the boundaries between physical and virtual (or computing) worlds are becoming transparent with the possibilities of seamless integration of the two worlds. According to (Aarts and Wichert, 2009, Aarts and Encarnação, 2008) ambient intelligence refers to “the vision of integrating computational intelligence within human environments and the artifacts that it contains such that human-centered services are offered to satisfy human agents’ immediate needs.” Such environments are expected to offer services that are centered on the human agents within the environment, i.e. services that are adaptive and personalized to the human agents. Satisfying a human agent involves not only providing the functionalities, but to understand and adapt such functionalities according to the human agent’s capabilities, limitations, likes, preferences, etc. Human-centered factors like perception, action, intention and attention described in chapter 3 play an important role in satisfying the human agent within such environments. Shifting the attention from a typical technology-oriented research in ambient intelligence to a more human-
centered and experience-oriented research, introduces a need to develop novel concepts in developing an infrastructure for ambient intelligence. Egocentric interaction (refer to chapter 3) offers such a conceptual platform for (re-)designing and evaluating infrastructures for ambient intelligence.

4.1.1 Ambient Ecology

Ambient ecology (Alvarez et al., 2006, Kameas et al., 2009) can be considered as the infrastructure through which ambient intelligence can be realized. It refers to an inter-connected collection of heterogeneous components like smart objects (Fujinami et al., 2005), middleware components, ambient intelligence applications, virtual objects that are part of such applications and are accessed through the smart objects, artificial agents, and human agent(s) with a common goal of supporting human agents’ lifestyle and well-being. Ambient ecologies provide a platform for supporting everyday human activities through ambient intelligence applications. The computers within an ambient ecology are expected to be both physically and cognitively invisible, embedded within everyday objects, and stay in the background allowing human agents to perform foreground activities (Weiser and Brown, 1996, Ishii and Ullmer, 1997). This fundamental change in the view of computers introduces a need to explore novel interaction paradigms (such as egocentric interaction explored in this thesis) that are human and their activity-centered, instead of being device-centered as with the WIMP (windows, icons, menus and pointing devices) interaction paradigm for desktop computers. An ambient ecology should be context-aware (Dey, 2001, Schmidt, 2002) since they exist in everyday human environments, and should allow human agents to provide implicit input (Schmidt, 2000) to the ambient intelligence applications relaxing the need for explicit attention during all the interactions within the ecology.

This thesis attempts to blend the physical and the virtual aspects of an ambient ecology by taking a physical-virtual design perspective in supporting physical-virtual activities (Pederson, 2003). An ambient ecology is expected to be a sensitive, adaptive, and responsive environment with regard to its physical and virtual aspects. It represents a physical-virtual space of activity possibilities and a set of constraints with an overall goal of fulfilling the individual occupant’s immediate needs. Human agents within an ambient ecology possess dual citizenship, one in the physical world and the other in the virtual world. The term environment refers to the surrounding in which a human agent is situated (both physically and virtually as discussed in chapter 3). An ambient ecology offers the infrastructure for easy movements across the physical-virtual boundary, thereby reaping the best of both realms. The sensitive, adaptive and responsive nature of an ambient ecology is dependent on a human agent’s ability to perceive and potentially act at a particular moment in time, thereby defining intelligence with reference to the human agent. The situative space model described in chapter 3 is useful not only in modeling a human agent’s situation within an ambient ecology, but also in recognizing their activities and actions (refer to chapter 8), and in facilitating
The easy ADL Ecology: an Infrastructure for Ambient Intelligence

4.2 The easy ADL Ecology

The easy ADL ecology is an infrastructure for realizing the vision of ambient intelligence, designed and developed based on the principles of egocentric interaction. The development of the easy ADL ecology began as part of the easyADL project (easyADL – Independent life despite dementia) (Backman, 2006) funded from the EC Target 1 structural fund program for Northern Norrland, Sweden. The easyADL project attempted to address age-related healthcare challenges which is becoming a significant problem for individuals (like the caregivers and the relatives), and to the society from a humane and economical perspective. The easy ADL ecology is envisioned as a “cognitive prosthesis” to facilitate older citizens (especially mild-dementia patients) to lead an independent life at home performing their activities of daily living (ADL) instead of having to move to healthcare institutions. The easy ADL ecology follows a three-step development methodology: initially, an immersive virtual reality simulated home environment (Surie and Pederson, 2007) was used for the development, followed by a living laboratory physical home environment based development. In the future, the aim is to develop and deploy the easy ADL ecology for in-situ environments, and to support ambient assisted living (AAL) with the inclusion of assessment of health conditions of human agents within the ambient ecology (Lindgren et al., 2011).

4.2.1 Immersive Virtual Reality Home Environment

Using virtual reality for simulating a smart home is not new and is used in many research works including UbiREAL (Nishikawa et al., 2006), eHomeSimulator (Ibrahim and Daniel, 2007), TATUS (O’Neill et al., 2005), CASS (Park et al., 2007), and C@sa (De Carolis et al., 2005). The immersive virtual reality simulated home environment included a VR model that was developed using the Colosseum3D real-time physics platform (Backman, 2005). The VR simulation-based approach facilitated the simulation of sensors, actuators, interfaces and interactive devices, physical objects and virtual objects, thereby allowing the focus to be on the conceptual and the algorithmic parts of the easy ADL ecology, ignoring the technological aspects (Sjölie, 2007). Concepts belonging to egocentric interaction are explorative, and a simulation environment was ideal during the initial development phase to lower the cost of (re-) designing various aspects of the easy ADL ecology, especially since physical implementation can be ignored and the iterative process of concept development, design, implementation and evaluation can be performed with ease.

User experience evaluation of the easy ADL ecology is important since ambient intelligence technologies are built to be human-centric (Abowd and Mynatt, 2005). The VR simulation was useful for rapid prototyping and performing human-centric evaluations almost throughout the development.
**Fig. 4.1.a.** The easy ADL ecology simulated in an immersive virtual reality home environment.

**Fig. 4.1.b.** The easy ADL ecology simulated in an immersive virtual reality home environment (Surie and Pederson, 2007).
process. The VR-based approach does have limitations with the setup: for example, the human agent’s mobility is limited by the sensing range of the magnetic tracker, and there is a lack of physical affordances while interacting with objects.

4.2.2 Living Laboratory Physical Home Environment

The easy ADL ecology was physically developed as a living laboratory for ambient intelligence research initially within the University campus and then moved to an apartment in Umeå, Sweden. The living laboratory environment facilitated the possibilities to develop and evaluate technologies that are dependent on physical space. Also, a living laboratory setup allows for conducting authentic yet adaptable and controllable experiments (Kidd et al., 1999, Intille et al., 2005) that are important during the initial stages of prototype development. Aspects that need to be evaluated and improved can be focused at a cheaper cost in comparison to doing such experiments directly in in-situ environments. The VR-based simulation approach and the living laboratory approach complement each other focusing on different aspects of the easy ADL ecology. For instance, networking, tracking and localization of smart objects could be better experimented with in a physical living laboratory home environment, while developing activity recognition algorithms using simulated sensors could be experimented with better in a VR environment. The living laboratory home environment is a 54 m² apartment intended for providing ADL support for single occupants (other occupants could still be a part of the easy ADL ecology without receiving computing support). The restriction of supporting single occupants allowed for conducting simple experiments and avoided the need to address challenges inherent in multi-occupant environments considering the scope of this thesis. However in the future, the aim is to transform the easy ADL ecology into a home for multi-occupants receiving computing support since ambient ecologies are usually occupied by multiple human agents. The Bremen Ambient Assisted Living Lab is intended to provide mobility support for elderly and people with physical and cognitive impairments (Bremen_AAL_Lab, 2011) within a functional bathroom and kitchen environment for two persons. LIVING LAB (LivingLab, 2011) is a research and development infrastructure intended to explore human interaction and innovations surrounding smart homes. The LIVING LAB is stationed in several houses across Europe providing cross-cultural and cross-climatic research possibilities.

4.2.3 Infrastructure of the easy ADL Ecology: Overview

The easy ADL ecology is a complex system comprising many entities including smart objects (both environmental and wearable objects), computing devices and components, and a human agent in the middle of all. Many of these entities possess individual properties and behave with varying levels of autonomy within the easy ADL ecology. Together, they form an ecology where
they communicate, co-operate and interact with other entities with the overall goal of providing a positive experience to the human agent. The environmental smart objects communicate with the wearable computer using ZigBee and/or WLAN protocols, while the wearable smart objects use Bluetooth protocol and/or connected using wires. Refer to Fig. 4.2. Ambient intelligence applications run on top of the middleware in a centralized manner and synchronized copies of those ambient intelligence applications run on selected smart objects (depending on the smart object) supporting the human agent with their everyday physical-virtual activities. The technical and technological challenges in developing the easy ADL ecology like tracking the state changes to physical and virtual objects (refer to chapter 6), situative space tracking of the physical and the virtual objects (refer to chapter 7), and activity and action recognition (refer to chapter 8) are also addressed in this thesis.

The easy ADL ecology is not restricted in supporting physical activities as is often the case in similar research. Occupants of a home often perform everyday virtual activities like checking emails, paying monthly bills, chatting online, etc., and by introducing an infrastructure that attempts to integrate the physical and virtual aspects of an environment, the occupants are more likely to perform a third class of activities referred to as physical-virtual activities (Pederson, 2003).

4.3 Smart Objects

Smart objects are ordinary physical objects with additional capabilities that provide complementary virtual (or computing) services without compromising on their established primary purpose (Kawsar et al., 2008) as described in chapter 2. Virtual services include sensing and recognizing their states and the surrounding context, communicating information with other smart objects and the personal activity-centric middleware, facilitating actuation and human interaction with ambient intelligence applications, etc. In providing virtual services, smart objects are expected not to compromise on their appearance and their interaction metaphor which is an interesting and largely unexplored challenge. Aesthetics and other artistic factors that are not technical by any means come into play in designing smart objects. Such an exploration requires user experience evaluation in (re-) designing and building smart objects described in chapter 5.

Smart objects are expected to assist and support human agents with their activities within an ambient ecology and are important building blocks in addressing ambient ecology challenges like recognizing human activities (addressed in chapter 8), localizing objects and human agents within the ambient ecology (addressed in chapter 7), and facilitating human interaction within ambient intelligence applications (addressed later in this chapter).
Smart objects possess varied computational and technological capabilities. At one end there are everyday objects that are tagged with passive tags where the smart object's intelligence is located in an external infrastructure, at the other end there are everyday objects with built-in intelligence able to act autonomously like intelligent robots. In this thesis, smart objects are classified depending on the type of mediators that it is augmented with. The type of mediators to a large extend determines the underlying technologies.
used in building those smart objects. Smart objects augmented with interactive mediators like LCD screens, audio speakers, microphones, touch screens, etc. communicate with the wearable computer using WLAN technology, forming a local area network for media exchanges. Refer to Fig. 4.3:

- **Fig. 4.3. a & b) Bedroom photo frame** smart object that usually displays photos of relatives, friends, nature, fantasy, etc. depending on the human agent’s context. It also provides access to other virtual objects like alarm clock, day schedule, weather information, dressing assistant, etc. depending on the context.

- **Fig. 4.3. c & d) Bathroom mirror** smart object that usually presents latest news, personal information like day schedule, emails, etc. in parallel to reflecting the human agent’s face.

- **Fig. 4.3. e & f) Refrigerator** smart object usually provides access to virtual objects that keep track of the expiry date of food items within the refrigerator, suggests recipes depending on the ingredients available within the refrigerator, etc. Virtual objects that are not associated to any particular physical object like bus timetable, news, emails, etc. are also presented and provide access to depending on the context.

- **Fig. 4.3. g & h) Showcase** smart object presents virtual objects like calendar, photos, etc.

- **Fig. 4.3. i & j) Cutting board** smart object provides access to virtual objects like recipes, cooking assistant, etc. in parallel to allowing the human agent to cut food items like vegetables and meat.

- **Fig. 4.3. k & l) Bookshelf** smart object stores books, CDs and DVDs, in parallel provides information about the latest books from a specific author available in the nearest book store, IMDB ratings and comments, and allow the human agent to access a virtual library from which books could be browsed, borrowed, ordered, etc.

- **Fig. 4.3. m & n) Dining table** smart object allows the human agent to have breakfast, lunch, fika, etc. in parallel to providing access to relevant virtual objects.

- **Fig. 4.3. o) Entrance door** smart object usually reminds the human agent to wear proper jackets, gloves, etc. depending on the weather information, welcomes the human agent when he/she comes home, etc.

- Smart objects augmented with interactive mediators present and provide access to virtual objects considering the human agent’s contextual conditions like their activity and action context; situational context described using the situative space model; etc.
and the needs expressed by the ambient intelligence applications. The interaction management rules to be described in section 4.6 is used in determining if, when, where and how to present the virtual objects on smart objects in the easy ADL ecology.

Fig. 4.3. Smart objects augmented with interactive mediators providing access to virtual objects within the easy ADL ecology.
Smart objects augmented with sensor motes communicate with the wearable computer using ZigBee technology forming a wireless sensor network sufficient for low data rate communication in the easy ADL ecology. Refer to Fig. 4.4:

- Fig. 4.4. a) Refrigerator smart object with embedded temperature sensors, light sensors, etc. presents information like the internal temperature of the refrigerator and the freezer, if their doors are open or closed, etc.

- Fig. 4.4. b) Microwave oven smart object with embedded pressure sensitive pad, on-off switches, light sensor, rotation sensors, etc. that allow for capturing the internal states of the microwave oven and also the state changes caused by a human agent's manipulation.

- Fig. 4.4. c) Hand wash smart object that detects when it is used.

- Fig. 4.4. d) Coffee maker smart object that detects the quantity of coffee beans in it and the type of coffee selected by a human agent.

- Fig. 4.4. e & g) Stove smart object that detects if the stove and/or the oven is on or off, the stove's temperatures at different heating plates, the oven's temperature, etc.

- Fig. 4.4. f) Waste bin smart object that detects if it is full and needs to be emptied.

- Fig. 4.4. h) Dish washer smart object that detects if it is on or off, the current program selected, if the dishwasher is open or closed, etc.

- Fig. 4.4. i) Cutlery drawer smart object that detects if it is open or closed.

- Fig. 4.4. j) Toilet closet smart object that detects when the closet is flushed.

Smart objects that are wearable and communicate with the wearable computer using Bluetooth technology and/or wired connection form a body area network in the ambient ecology. Refer to Fig. 4.5. The headset smart object facilitates speech based interaction with virtual objects in the ambient ecology, while the bracelet smart object on each hand provide acceleration values along the 3 axis useful in facilitating gesture based interaction with virtual objects. For more information about speech recognition, gesture recognition and the accuracy of causing state changes to virtual objects, refer to chapter 6.
The aim is to reduce the heterogeneity amongst the smart objects within the easy ADL ecology by channeling the interaction between the human agent and the objects through an activity-centric middleware running on a wearable computer (refer to section 4.4). The middleware is built on several of the principles of egocentric interaction like situatedness, the physical-virtual equity principle, the proximity principle, replacing input and output by perception and action, etc. presented earlier in chapter 3, and also inspired by Intel’s personal server concept (Want et al., 2002) where a wearable computer allows a human agent to readily store and process virtual objects while providing access to those virtual objects through the interfaces found in the smart objects. The design choice of centralizing as much of the computation and sensing to a couple of wearable device(s) instead of embedding it all in the smart objects is due to an expected reduction in design complexity (less attention needs to be directed towards mitigating the problem of uneven conditioning (Satyanarayanan, 2001), increased privacy, and reduced cost.

Fig. 4.4. Smart objects augmented with sensor motes within the easy ADL ecology (Surie et al., 2008).
The heavy calculations are performed by the wearable computer instead of expecting the smart objects to perform such calculations. There exists a continuum between purely wearable computing approaches to approaches that are purely distributed through co-operative smarts objects within an ambient ecology (Rhodes et al., 1999). The approach presented in this thesis lies somewhere in between these two extremes, but more closer to the wearable computing approach.

![Image](image_url)

**Fig. 4.5.** A *bathroom mirror* smart object presenting day schedule and activity reminder virtual objects manipulated through speech and gestures using a *headset* smart object and the *bracelet* smart objects worn on both the hands by a human agent. The *bathroom mirror* smart object reflects the human agent’s face in addition to presenting the virtual services.

In an ambient ecology inhabited by multiple human agents (undoubtedly very common), the inherent complexity of individual smart objects might be higher than the ones described in this thesis. It is deliberately chosen to start exploring egocentric interaction using the simple case of one single human agent leaving potential extensions towards explicit support for collaborative human activities for future work.
4.3.2 Entity-Relationship Representation of Smart Objects

The entity-relationship (E-R) diagram for smart objects as modeled by the personal activity-centric middleware (refer to section 4.4) is presented in Fig. 4.6. The E-R diagram represents many of the principles of egocentric interaction like situatedness, the physical-virtual equity principle, replacing input and output by perception and action, etc. described in chapter 3. Smart objects that are wearable and the ones that are in the environment are treated alike in their representation even though there are inherent differences in their properties, usability, etc. Their differences are known (based on their identity information) and used by the personal activity-centric middleware during its decision making process.

In the E-R diagram, a smart object is considered as a physical object that provides access to virtual objects through the mediators present in it. Physical objects, virtual objects and mediators within the easy ADL ecology are entities that are modeled based on the physical-virtual equity principle (refer to chapter 3) supporting the concepts of physical-virtual artefacts, physical-virtual situations and physical-virtual activities (Pederson, 2003) as shown in Fig. 4.6. |p|, |m| and |v| refer to the cardinality of the relationship between physical objects, mediators and virtual objects associated with a smart object, respectively. For a smart object, |p| is always 1 since the smart object has to be a physical object. A purely virtual object is not considered to be a smart object even though it might offer smart services without being associated to specific physical objects (or) mediators in the easy ADL ecology. A smart object should contain at least one mediator and |m| can be greater than or equal to 1. If |m| is equal to 0 then such objects are referred to as “plain” objects requiring external mediators to enable them to be a part of the ambient ecology. A smart object may in some cases not provide access to virtual objects leading to |v| = 0 at times.

In the implemented ambient ecology, a smart object has an object identity ObjID and an object type ObjType. A smart object is a physical object with characteristics and functionalities of the physical object dependent on the physical object type PhyObjType. For instance, a refrigerator smart object may represent a refrigerator physical object with a freezer or without a freezer depending on PhyObjType. The refrigerator physical object has PhyObj internal state attributes including temperature, on-off, etc. while the LCD display mediator has mediator state attributes like display brightness, screen size, etc. The loudspeaker mediator has mediator state attributes like loudness, on-off, etc. Virtual objects are dependent on the mediators (usually interactive devices) to enable human agents to access them and interact with them. Physical objects on the other hand, depend on mediators (usually sensors) for allowing the personal activity-centric middleware to keep track of their internal states and external states (i.e., states with reference to other physical objects in a physical space).
When a human agent moves around and acts in physical space, the situative space model associated with them is subject to dynamic changes. Smart objects including mediators enter and/or leave the situative spaces and sets as an effect. Smart objects that are wearable are usually more stable within the situative spaces and they automatically become a part of the situative spaces the moment a human agent wears them. In environments with a limited amount of stationary (embedded) mediators, wearable mediators are useful in making sure that the human agent is pervasively connected to the virtual world, reducing the problem of uneven conditioning (Satyanarayanan, 2001).

$|po|$, $|me|$ and $|vo|$ refer to the cardinality of a particular set of physical objects, mediators and virtual objects that are associated with each other to form physical-virtual artefacts (PVA) (Pederson, 2003). $|PVA|$ refers to the

---

**Fig. 4.6.** Entity-relationship diagram of a smart object within the easy ADL ecology.
set of physical-virtual artefacts currently known to the personal activity-centric middleware. Note that the associative bonds between physical objects, virtual objects and mediators that form PVAs can be both strong and permanent (e.g. created by the designer of the PVA) or weak and temporary (e.g. created dynamically by the activity-centric middleware depending on the human agent’s context). Also, note that there is just one situative space model (refer to chapter 3) for each individual human agent within the easy ADL ecology and all the physical objects, virtual objects and mediators known to the personal activity-centric middleware are represented in the model. Physical objects, virtual objects and mediators play a role in physical-virtual activities (Pederson, 2003) and are associated to human activities and actions as described in the E-R diagram. Since the smart objects that are part of the easy ADL ecology are managed by the personal activity-centric middleware, they function in an activity-centric manner while offering services to the human agent. Human activities are regarded to be in one of several possible states, namely initiated, interrupted, resumed and completed. The activity recognizer to be described in Section 4.4.3 keeps track of these activity states.

4.4 A Personal Activity-Centric Middleware

A middleware is useful for handling the heterogeneity of smart objects; it is also useful for dealing with other sources of complexity involved in ambient intelligence applications by encapsulating the low-level sensing and networking tasks, context recognition, and supporting human interactions. The personal activity-centric middleware is intended to facilitate the handling of human-centered parameters inspired by egocentric interaction such as perception, action, intention and attention in a unified manner.

Considering the number and diversity of smart objects in the easy ADL ecology, the middleware has to address many of the requirements of traditional distributed systems such as interoperability, scalability, security, and tolerance for component failures and disconnections (Henricksen et al., 2005). Egocentric interaction introduces additional requirements including the modeling and tracking of the smart objects, the situative spaces, and the agent’s activities and actions – information which is used to facilitate human interaction with the ambient intelligence applications.

The personal activity-centric middleware is composed of four components, namely the object manager, the situation monitor, the activity recognizer, and the interaction manager, as shown in Fig. 4.7. The ambient intelligence applications run on top of the middleware, and their synchronized copies run on specific smart objects to reduce the amount data communication between the wearable computer and the smart objects with some processing done locally on the smart objects. The middleware components run as independent modules and offer a common interface to communicate with the ambient intelligence applications. Microsoft Message Queues (MSMQ) (Redkar, 2004) was used for exchanging data objects and strings among the middleware components, the ambient intelligence applications and the smart objects. It is important to keep down the number of queues since each queue added
introduces additional overhead in exchanging messages. Instead of one queue for each individual communication path, which also might reduce flexibility when new components are added, a single queue for each individual component is used.

The queues work like a mailbox where sent messages are collected and can be read periodically. During the initialization phase, the existing queues are emptied and newly arriving messages are listened to. The middleware is implemented in C# while some modules of the object manager are implemented in C++ and some modules of the activity recognizer in Matlab. In the future, the aim is to convert the code written in Matlab to C#.

4.4.1 Object Manager

The object manager is responsible for the following:

- To manage the smart objects within the easy ADL ecology.
- To initiate and manage wireless communication (and wired communication) with smart objects.
- When a new smart object is detected in the ambient ecology, the object manager attempts to obtain more information about the smart object from its manufacturer’s database (Kulkarni et al., 2006). A mock-up database is used with an assumption that in the future, smart object manufacturers would maintain such a database online considering the interest shown by the industry and academia in the Internet of Things (Gershenfeld et al., 2004).

Since smart objects usually contain physical objects, virtual objects and/or mediators, real-time information about them are also maintained in an object-centric manner by the object manager. Maintaining information includes not only to identify the objects, but also to store up-to-date information about their attributes and current values. For instance, information about an object’s association to relevant activities and actions (from the activity recognizer), the object’s position in the situative spaces at a particular moment in time (from the situation monitor), an object’s current internal states, external states, etc., are maintained. Refer to Fig. 4.6 for the E-R diagram.

- The object manager keeps track of the physical objects and mediators by sensing and communicating with the smart objects in the ambient ecology. It keeps track of the virtual objects by communicating with the interaction manager and the ambient intelligence applications running on the smart objects.
- The object manager keeps track of the symbolic location of physical objects and mediators (like “in the bathroom”, “in the kitchen”, etc.) by associating the individual objects to co-located stationary objects with known symbolic location. Since indoor environments contain walls that often act as a boundary for the perception and action spaces, objects within these spaces often (but not always) can be located to a specific room.
The easy ADL Ecology: an Infrastructure for Ambient Intelligence

Fig. 4.7. An infrastructure for the easy ADL ecology comprising of smart objects, a personal activity-centric middleware, ambient intelligence applications and a human agent in the middle of it all.
• To exchange information with other middleware components and offer them services related to smart objects. For instance, the interaction manager presents and provides access to virtual objects on smart objects by the mediation of the object manager.
• To exchange information with ambient intelligence applications running on top of the middleware and offer them services related to smart objects. The ambient intelligence applications are managed in a centralized manner by the middleware and they cannot directly interact with the smart objects. To maintain and exchange information about the physical-virtual artifacts formed by the association of physical objects, virtual objects and mediators.

4.4.2 Situation Monitor

The situation monitor is responsible for the following:
• To keep track of the identity and location of the physical objects, virtual objects and mediators within the human agent’s situative spaces (refer to chapter 3) by communicating with the object manager and the smart objects mediated by the object manager. Also, to keep track of the states and state changes of these objects (important in keeping the examinable set and the manipulated set updated).
• To exchange information with other middleware components and offer them services related to the agent’s situation. For instance, the interaction manager can decide if, when, where and how to position virtual objects within the individual situative spaces by getting up-to-date information about the content of the situative spaces while the activity recognizer recognizes activities and actions using the content of the situative spaces (Surie et al., 2007b).
• To exchange information with applications running on top of the middleware and offer them services related to the agent’s situation, in effect making them situation-aware without having to compute the situation locally within the applications.

The primary approach taken for situative space tracking is to use wireless LAN signal-strength-based localization as described in chapter 7. ZigBee-based wireless sensor networking of smart objects is used to keep track of the state changes (examinable set and manipulated set) to physical objects and mediators as described in chapter 6.

4.4.3 Activity Recognizer

The activity recognizer is responsible for the following:
• To model and recognize the agent’s activities and actions using information channels provided by the situation manager. The different information channels contain information about the content of the different situative spaces and sets over time. Two different activity
recognition systems have been built based on these information channels, for details refer to chapter 8.

- Modeling human activities and actions are done offline (i.e. while not performing the task of activity and action recognition) and in a supervised manner during an explicit training phase. In a future implementation, the aim is to model human activities using online learning algorithms. Supervised learning is preferred to unsupervised learning because of the complexities involved in modeling and recognizing everyday human activities and actions. Supervised learning allows for personalized modeling of human activities and actions, and facilitates their recognition with higher accuracy.

- To derive activity-centric information including: a) the set of objects that are associated to individual activities and actions; b) the set of actions and operations that are mandatory for individual activities; and c) important events that determine the initiation, interruption, resumption and completion of individual activities.

- To exchange information with other middleware components and offer them services related to human activities and actions. For instance, the interaction manager communicates with the activity recognizer to know the human agent’s current activity and action context for deciding if, when, where and how to present and provide access to virtual objects within the ambient ecology. Other services include providing activity-centric information to the object manager that uses this information in associating individual objects known to the object manager with their respective activities and actions.

- To exchange information with ambient intelligence applications and offer them services related to human activities and actions. For instance, activity-centric information and services are useful for applications like the cognitive prosthesis that attempts to provide activity support for mild-dementia patients within an ambient ecology (Backman, 2006).

### 4.4.4 Interaction Manager

The interaction manager is responsible for the following:

- To facilitate interaction with virtual objects (part of the ambient intelligence applications) based on the principles of egocentric interaction and the situative space model (refer to chapter 3). Even though the focus is on facilitating interaction with virtual objects, the interaction manager handles both physical and virtual objects, physical and virtual situations, and physical and virtual activities from an egalitarian perspective as per the physical-virtual equity principle.

- The interaction manager facilitates situated interaction by considering the current content of the situative spaces (from the situation monitor) and making decisions concerning if, when, where and how to present virtual objects so that they can be perceived and/or acted upon by the human agent.

- The interaction manager facilitates activity-centered interaction by considering the human agent’s activity and action context (from the
activity recognizer), and the virtual objects' association to the current activity and action context (from the object manager).

- The interaction manager facilitates multimodal interaction by considering the mediators available within the situative spaces (from the situation monitor), their current states (from the object manager) and the virtual objects' mediator and modality preferences (information from the ambient intelligence applications). Multimodal interaction improves: (a) the bandwidth of interaction both in terms of quantity and quality; and (b) the usability by offsetting the weakness of one modality by the strengths of another in various contexts.

- The interaction manager facilitates mixed-initiative interaction by allowing both the human agent and the ambient intelligence applications to initiate interaction sessions. The interaction manager uses separate sets of interaction management rules (refer to section 4.6) for handling the virtual objects in human-initiated interaction sessions and ambient intelligence application-initiated interaction sessions. Mixed-initiative interaction enables the agent to exert control in parallel with proactively operating ambient intelligence applications in the background to better satisfy the agent's computing needs.

- The interaction manager facilitates ambient interaction by considering the human agent's peripheral and central attention capabilities, and the virtual objects' attention requirements (by communicating with the ambient intelligence applications). Human attention is modeled at two levels of abstraction: a) at a low-level, human attention is driven by objects within the situative spaces (obtained by communicating with the situation monitor); and b) at a high-level, human attention is driven by human intentions (obtained by communicating with the activity recognizer).

- To handle additional virtual object attributes like privacy, importance, etc. that are important to be considered and is established by the ambient intelligence applications. Privacy is addressed by considering the symbolic location of the mediators in the ambient ecology and augmented within smart objects.

- To exchange information with other middleware components and offer them services related to human interaction with virtual objects within an ambient ecology.

- To exchange information with ambient intelligence applications and offer them services related to human interaction with virtual objects in the ambient ecology.

4.5 Ambient Intelligence Applications

Ambient intelligence applications run on top of the personal activity-centric middleware and synchronized copies run on selected smart objects for the reason mentioned earlier. The applications on the smart objects are synchronized, managed and controlled by the middleware in a centralized manner. Ambient intelligence applications communicate with other middleware components to obtain relevant contextual information about the
easy ADL ecology including the human agent in it. The applications contain domain specific knowledge and by using the additional contextual information, they become context-aware thereby satisfying the general requirements for being a part of an ambient ecology. Ambient intelligence applications communicate with the human agent by presenting virtual objects that are filtered and managed by the interaction manager according to the interaction management rules to be described in section 4.6. These virtual objects maintained by the ambient intelligence applications possess many attributes that are relevant within the ambient ecology (e.g. activity association, attention load, associated smart objects, etc.) which traditional virtual objects belonging to typical desktop applications (accessed through mouse, keyboard and a screen) do not possess.

In the specific design case reported in this thesis, the ambient intelligence applications have been designed to provide physical-virtual activity (Pederson, 2003) support to people in the early stages of dementia. The easy ADL ecology (where ADL stands for activities of daily living as commonly understood within the medical field) is intended to provide human agents with independence and satisfaction to an extent that a normal non-augmented environment cannot. The research effort is long-term and the work presented here includes 19 mock-up (limited functionality and/or partially implemented) applications introducing more than 100 virtual objects used for an initial user experience evaluation (presented in chapter 5) of the easy ADL ecology as such. Applications include medicine reminder, old-food-in-the-refrigerator reminder, food recipe provider, diet controller, shopping assistant, safety and security manager, weather information provider, clothing assistant, news provider, transportation information provider, day scheduler, etc.

### 4.6 Interaction Management Rules

The interaction manager provides access to virtual objects within the easy ADL ecology on the basis of interaction management rules that answer the important questions of *if*, *when*, *where* and *how* a virtual object should be made present and accessible on request by a) the human agent or b) an ambient intelligence application. The rules are prioritized according to which question they address with the *if* question having the highest priority, followed by the *when* question, the *where* question, and the *how* question.

In Table 4.1, *if* question is linked to the virtual object attributes *importance*, *activity association*, and *action association*. Similarly, the *when* question is associated with *session duration* and *session timeout*; the *where* question to *attention load*, *associated smart objects*, and *privacy*; and (finally) the *how* question to *modality preference*. The interaction techniques described in section 4.7 also address the *how* question.

The virtual objects in the interaction manager either enter a *human-initiated interaction session* or the *application-initiated interaction session* depending on who initiated interaction with different sets of interaction
management rules. The interaction manager filters the virtual objects and inserts only those virtual objects that satisfy the necessary conditions of the interaction management rules into the dispatching queue. The dispatching queue is a simple priority queue that constantly gets updated (every 2 to 3 seconds) with a set of virtual objects depending on the contextual conditions within the ambient ecology and the current needs of the ambient intelligence applications. The update rate is dependent on the low-level sensing and networking infrastructure of the easy ADL ecology (which could be improved with technological advancements in the future), and some of the personal activity-centric middleware components like the activity recognizer and the situation monitor. The update rate is not an issue since human agents usually take a few seconds to stabilize themselves in a situation before virtual objects could usefully be presented to them. Handling conflicts among the interaction management rules is a major challenge and in the event of a conflict, the virtual objects part of a human-initiated interaction session is given priority over other virtual objects part of an AmI application-initiated session.

It should be noted that several sessions usually run at the same time. Within a session, the *importance* attribute is given the highest priority followed by *activity association* and *action association*, *session duration*, *session timeout*, etc. *Modality preference* is given the least priority. Refer to Table 4.1. These interaction session rules are based on egocentric interaction principles and may be altered in future versions of the system as the ambient ecology setup evolves. For instance, the priorities could be personalized to a human agent both implicitly by learning and modeling their interaction experience, and explicitly by allowing the human agent to modify the priorities.

Refer to Table 4.1. The importance attribute value of a virtual object may change while being a part of an interaction session. For instance, a virtual object with low importance can suddenly move to medium importance and/or high importance depending on the context and the passage of time. The virtual object’s attribute values are handled by the respective ambient intelligence applications and updated instantly before the virtual object is filtered by the interaction manager. Some virtual objects are associated to specific activity and action context. Such objects are presented by taking into consideration a human agent’s current activity and action context. Also, note that some virtual objects are not dependent on the human agent’s current activity and action context. Such virtual objects are handled by associating them to all activity and action contexts modeled within the ambient ecology before applying the interaction management rules described in Table 4.1. Similarly, some virtual objects are not dependent on specific smart objects available in the ambient ecology for being presented and such virtual objects are considered to be associated to all smart objects in the ambient ecology.
Table 4.1. Interaction management rules within the easy ADL ecology.

<table>
<thead>
<tr>
<th>Virtual Object Attributes (values)</th>
<th>Human-Initiated Interaction Session</th>
<th>Application-Initiated Interaction Session</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Initiation: Selection of a virtual object by a human agent.</td>
<td>Initiation: Selection of a virtual object by an application.</td>
</tr>
<tr>
<td></td>
<td>Termination: De-selection of a virtual object by a human agent.</td>
<td>Termination: Virtual object reaches the session timeout or the end of its session duration.</td>
</tr>
<tr>
<td>Importance (high, medium, low)</td>
<td>Importance value ignored. Virtual object is presented immediately.</td>
<td>Importance value is considered.</td>
</tr>
<tr>
<td></td>
<td>High: Virtual object is presented immediately.</td>
<td>High: Virtual object is presented immediately.</td>
</tr>
<tr>
<td></td>
<td>Medium: Virtual object is presented if no high importance virtual objects are in the dispatching queue.</td>
<td>Medium: Virtual object is presented if no high importance virtual objects are in the dispatching queue.</td>
</tr>
<tr>
<td></td>
<td>Low: Virtual object is presented if no high and medium importance virtual objects are in the dispatching queue.</td>
<td>Low: Virtual object is presented if no high and medium importance virtual objects are in the dispatching queue.</td>
</tr>
<tr>
<td>Activity association (set of activities)</td>
<td>Activity and action association ignored. Virtual object is presented immediately.</td>
<td>Activity and action association is considered.</td>
</tr>
<tr>
<td></td>
<td>Activity and action context matches: Virtual object is presented immediately.</td>
<td>Activity and action context matches: Virtual object is presented immediately.</td>
</tr>
<tr>
<td></td>
<td>Activity context alone matches: Virtual object waits in the dispatching queue for the action context to match.</td>
<td>Activity context alone matches: Virtual object waits in the dispatching queue for the action context to match.</td>
</tr>
<tr>
<td></td>
<td>Activity and action context does not match: Virtual object waits in the dispatching queue for the activity and action context to match.</td>
<td>Activity and action context does not match: Virtual object waits in the dispatching queue for the activity and action context to match.</td>
</tr>
<tr>
<td>Session duration (duration in seconds)</td>
<td>Session duration value is ignored. Virtual object presentation continues until the human agent terminates it.</td>
<td>Session duration value is considered.</td>
</tr>
<tr>
<td></td>
<td>Session duration is not reached: Virtual object presentation continues.</td>
<td>Session duration is not reached: Virtual object presentation continues.</td>
</tr>
<tr>
<td></td>
<td>Session duration is reached: Virtual object presentation terminates.</td>
<td>Session duration is reached: Virtual object presentation terminates.</td>
</tr>
<tr>
<td>Session timeout (timeout in seconds)</td>
<td>Session timeout value is ignored. Virtual object waits in the dispatching queue until the human agent terminates it.</td>
<td>Session timeout value is considered.</td>
</tr>
<tr>
<td></td>
<td>Session timeout is not reached: Virtual object waits in the dispatching queue.</td>
<td>Session timeout is not reached: Virtual object waits in the dispatching queue.</td>
</tr>
<tr>
<td></td>
<td>Session timeout is reached: The ambient intelligence application is informed about the failure in presenting the virtual object.</td>
<td>Session timeout is reached: The ambient intelligence application is informed about the failure in presenting the virtual object.</td>
</tr>
<tr>
<td>Attention load (peripheral attention, central attention, demand agent response)</td>
<td>Same mediator as selected by the human agent: Attention load is ignored. Virtual object is presented through that mediator.</td>
<td>Mediator is selected by the interaction manager: Attention load is considered.</td>
</tr>
<tr>
<td></td>
<td>Different mediator selected by the interaction manager: Attention load is considered. Rules 1, 2 and 3 apply.</td>
<td>Peripheral attention: Virtual object is presented if a peripheral mediator is available.</td>
</tr>
<tr>
<td></td>
<td>Central attention: Virtual object is presented if a central mediator is available.</td>
<td>Central attention: Virtual object is presented if a central mediator is available.</td>
</tr>
<tr>
<td></td>
<td>Demand agent response: Virtual object is presented if mediators capable of allowing the human agent to manipulate the virtual object is available.</td>
<td>Demand agent response: Virtual object is presented if mediators capable of allowing the human agent to manipulate the virtual object is available.</td>
</tr>
<tr>
<td>Associated smart objects (set of smart objects)</td>
<td>Same smart object selected by the human agent: Association to that smart object is ignored and the virtual object is presented.</td>
<td>Smart object is selected by the interaction manager: Association to that smart object is considered.</td>
</tr>
<tr>
<td></td>
<td>Different smart object selected by the interaction manager: Association to that smart object is considered. Rules 4 and 5 apply.</td>
<td>Associated: Virtual object is presented.</td>
</tr>
<tr>
<td></td>
<td>Not associated: Virtual object waits in the dispatching queue.</td>
<td>Not associated: Virtual object waits in the dispatching queue.</td>
</tr>
</tbody>
</table>
Note that when a human agent changes location or if the mediator selected by the human agent is not available for mediation anymore within human initiated interaction session, then the new mediator, the new smart object and/or the new modality for presenting and providing access to virtual objects are handled by the interaction manager. Then some of the virtual object’s attribute values like attention load, association to smart objects, privacy attribute, and modality preference which were previously ignored in this human initiated session are considered before presenting the virtual object in an appropriate manner within the situative spaces.

In Table 4.1, the interaction management rules consider the mediators that support demand agent response to also include mediators that support central and peripheral attention, while mediators that support central attention to also include mediators that support peripheral attention. Mediators that catch a human agent’s peripheral attention are referred to as peripheral mediators and mediators that demand central attention are referred to as central mediators in Table 4.1. Note that such a classification of mediators is done dynamically by the personal activity-centric middleware. Mediators that present personal virtual objects are referred to as personal mediators; mediators that present personal and private virtual objects are referred to as private mediators; and other mediators are referred to as public mediators.

Separate threads run to keep track of the virtual object’s session timeout attribute value and session duration attribute value continuously. A virtual object in the AmI application initiated interaction session can stay in the interaction manager’s dispatching queue, or can be continued to be presented

<table>
<thead>
<tr>
<th>Privacy (personal, private, public)</th>
<th>Same mediator selected by the human agent: Privacy is ignored and the virtual object is presented.</th>
<th>Mediator is selected by the interaction manager: Privacy is considered.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Different mediator selected by the interaction manager: Privacy is considered. Rules 6, 7, and 8 applies.</td>
<td><strong>Personal</strong> Virtual object is presented if a personal mediator is available.</td>
</tr>
<tr>
<td></td>
<td><strong>Private</strong> Virtual object is presented if a private mediator is available.</td>
<td><strong>Rule 7</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Public</strong> Virtual object is presented if a mediator is available.</td>
<td><strong>Rule 8</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Modality preference for access (visual, audio)</th>
<th>Same modality selected by the human agent: Modality preference is ignored and the virtual object is presented.</th>
<th>Modality is selected by the interaction manager: Modality preference is considered.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modality preference for manipulation (gesture, speech)</td>
<td>Different modality selected by the interaction manager: Modality preference is considered. Rules 9 and 10 applies.</td>
<td><strong>Preferred modality</strong> Virtual object is presented.</td>
</tr>
<tr>
<td></td>
<td><strong>Not the preferred modality</strong> Virtual object waits in the dispatching queue.</td>
<td><strong>Rule 10</strong></td>
</tr>
</tbody>
</table>

**Table 4.1 (continued).** Interaction management rules within the easy ADL ecology.
through smart objects as long as the virtual object’s session timeout or session duration is not expired, irrespective of the other contextual conditions.

4.7 Interaction Techniques

Egocentric interaction doesn't favor the use of any particular modality or mediator type for communication between the human agent and the surrounding ambient ecology. It only highlights the prerequisites for such communication to be at all possible: (1) mediators used for controlling or manipulating virtual objects/applications need to be reachable (inside the action space in the situative space model, Fig. 1) in the given situation; (2) mediators used for perceiving the presence and state of virtual objects need to be present themselves in the perception space of the given human agent. It is the role of the interaction manager (Fig. 4.7) to find the best combination of mediators given the situation. To increase the presence of mediators that enable manipulation of virtual objects, wearable mediators were relied on: accelerometer-fitted bracelets for identifying hand gesture commands; a headset microphone for speech command reception. For the presentation of the virtual objects, mediators on smart objects both in the environment (visual displays for visual presentation and loudspeakers for audial presentation) and the wearable ones (headphone for audial presentation) were used. Interaction techniques were developed which address the question of how to present and provide access to virtual objects in the ambient ecology. These interaction techniques were designed to accommodate for limitations of the technology that was at hand with future improvement possibilities. The overall goal has been to enable interaction with multiple objects at the same time over multiple interaction modalities distributed within the situative spaces.

Personalization of the situative spaces surrounding a human agent is an important challenge to address within an ambient ecology. Even though the problem is simplified in the context of this thesis where the scenario of multiple human agents within an ambient ecology is ignored, it is still important to identify the parts of the situative spaces that are personal and the ones that are anonymous. The selected and the manipulated sets are directly associated to object selections and manipulations respectively where the identity of the human agent causing such object manipulation are identified based on: the mediators worn by the agent (accelerometers for hand gesture recognition), or based on profile selection (speech recognition engine trained for specific user profiles). The action space and the perception space of a specific human agent are shared spaces that changes dynamically. Personalization of such spaces might form intersected volumes with other human agent's situative spaces introducing a need to handle shared personalization. Handling shared personalization and mechanisms to handle conflict of interests are left as future work. Similar to this challenge that exists at the space level, IdLenses (Schmidt et al., 2010) deal with personalization at the surface level where distinctive gestures are used to
personalize specific areas and hand contour analysis is used to identify the human agent.

Fig. 4.8. Gesture based manipulation of virtual objects part of the action space and in the easy ADL ecology.
4.7.1 Gesture based Object Manipulation

Gesture based manipulation enables a human agent to use hand gestures for selecting and manipulating virtual objects within the easy ADL ecology. Functionalities include (refer to Fig. 4.8):

- **Initiate-terminate gesture mode** where a human agent alters between using their hands normally in the physical environment to hand gestures for interacting with virtual objects. Note that when the gesture mode is initially, a set of virtual objects move to the action space (gesture modality) and when the gesture mode is terminated, those virtual objects are removed from the action space (gesture modality).

- **Mediator control** for selecting and deselecting mediators, and to changes their internal states. For example, gestures could be used to increase or decrease the volume of loudspeakers (mediator) equipped in a smart object as long as the mediator is present in the human agent’s action space.

- **Navigation (right, left, up, down, enter and exit)** for navigating among virtual objects within a smart object or application in the human agent’s action space.

- **Select and deselect** for selecting and deselecting virtual objects in the action space.

Gesture commands are also useful for associating specific gestures to human-defined virtual object manipulations similar to hot key commands of a keyboard. For example, gestures could be used to manipulate the “day schedule” virtual object directly instead of indirectly navigating through a visual display (equipped in a smart object) to first select the day schedule virtual object and then to manipulate it. This removes the need for compulsory object perception before object manipulation could be performed by a human agent.

4.7.2 Speech based Object Manipulation

Speech based manipulation enables a human agent to manipulate virtual objects in the action space using predefined speech commands. Functionalities include:

- **Initiate-terminate speech mode** where a human agent alters between normal speech (speaking with other human agents for instance) in the physical environment to speech intended to interact with virtual objects. Note that when the speech mode is initially, a set of virtual objects move to the action space (audial modality) and when the speech mode is terminated, those virtual objects are removed from the action space (audial modality).

- **Modality control** for selecting, deselecting and transforming the presentation of virtual objects from one modality to another in the action space.

- **Mediator control** for selecting, deselecting and changing the internal state of the mediators in the action space.
• Application control allows the human agent to initiate or terminate ambient intelligence applications.

Both the speech and the gesture commands use simple vocabularies and do not handle queries as in traditional gesture or speech based interaction languages. However, combining speech and gesture commands enhances the overall interaction vocabulary. For instance, a combination of speech and gesture commands allows a virtual object to be moved from the visual display of the Refrigerator smart object to the Dining Table smart object. The possibility for manipulating a specific virtual object through gestures and/or speech is dependent on the virtual object’s location within the situative spaces (should be in the action space). Refer to chapter 3.

4.7.3 Distributed Visual Presentation of Objects

Virtual objects are distributed among the mediators within the situative spaces enabling multiple virtual objects to be visually perceived by the human agent simultaneously. The spatial locations of the virtual objects (in turn depending on the location of their “host” mediator) determine if they can be visually perceived by the human agent through central attention or peripheral attention. Features of the interaction manager include:

• Ambient notification where a picture representing a virtual object changes its brightness to attract a human agent’s peripheral attention and present information as shown in Fig. 4.9.a. Association of the situative space model to human attention is described in section 3.6 where ambient notification belongs to peripheral attention.

• Virtual objects within the situative spaces. Virtual objects in the recognizable/examinable set attract central attention, while virtual objects outside them in the perception space attract peripheral attention. Virtual objects outside the perception space are not attended by the human agent for the moment. Refer to section 3.6.

• Size of the virtual object is altered in situations with only one visual mediator to move between central and peripheral attention as in the case with the bathroom mirror smart object in the bathroom.

• Multimedia visual presentation including text, icons, pictures, etc. depending on the visual virtual object.

• Visual feedback for virtual object manipulations.

• Mediator, modality and application statuses are visually presented as shown in Fig. 4.9.b.

4.7.4 Time-Multiplexed Audial Presentation of Objects

Audial presentation of virtual objects within the easy ADL ecology is time-multiplexed: only one virtual object is audially presented at a time. Once a virtual object begins to be audially presented, the next virtual object presentation is temporarily halted, waiting for the current audial
Fig. 4.9. Visual access to virtual objects provided by the dining table smart object where a human agent’s central and peripheral attention is considered depending on the location of dining table smart object within the situative spaces.
presentation to be completed, before it is presented. The location of the loudspeaker mediator within the situative spaces (if it is in the audial perception space or not) determines whether a virtual object can be potentially presented or not. Note that other object-related characteristics like object loudness, object type, etc. determine if the virtual object could actually be presented or not. Interaction manager features include:

- **Ambient notification** is used to present sounds and/or background music to represent virtual objects. The volume of ambient notification is usually lower in comparison to audible virtual objects needing a human agent’s central attention. Similar to visual modality, ambient notification in the audial modality belongs to peripheral attention as described in section 3.6.
- **Central attention** of the human agent is attracted by a beep sound with higher volume followed by the presentation of the audial virtual object. Refer to section 3.6.
- **Multimedia audial presentation** including speech, music, beep, etc. depending on the audial virtual object.
- **Audial feedback** for virtual object manipulation.
- **Mediator, modality and application statuses** are presented depending on the context.
Chapter 5

User Experience Evaluation of the easy ADL Ecology

This chapter describes a user experience (UX) evaluation of the easy ADL ecology within a living-laboratory home environment. The evaluation has been conducted to understand and explore the user experience issues, with some special attention to what might be achieved with respect to facilitating human interaction within the easy ADL ecology. The results are promising and give valuable input for further studies and development of the infrastructure, the ecology and the methodology, and ultimately the principles of egocentric interaction itself. The work reported here can be considered as an initial, yet concrete step towards the exploration and development of a unified human-centered interaction paradigm for ambient ecologies.

5.1 Introduction

To fully assess the value of using egocentric interaction as a tool for design of ambient ecologies is hard for several reasons. First of all, theoretically grounded design approaches for multi-device interaction systems which also incorporate everyday objects are still rare in the literature and thus a direct comparison is hard to make. While many system prototypes that fall into this category have been presented, more often than not the design process seem to be initiated and driven by the kind of devices and sensor/actuation technology that happened to be available. Available technology plays an important part also in egocentric interaction but it is not where this work starts. To adhere to the egocentric interaction design approach means to start with human perception and action capabilities, and to let everything else follow. Furthermore, explorative system design processes, such as the ones to which ambient ecology design belongs, are bound to be influenced by uncontrollable variables, making it very hard to conclude that a certain prototype implementation became better or worse than another due to the underlying design approach only. If design parameters were fewer or better controlled, a direct comparison of design approaches would have been possible. Like for the Ambient Intelligence community in general, ambient ecology designers still have some way to go until standardized experiments can be used as a way to compare design approaches and outcomes. Today, the developers and designers explore evaluation methods and metrics based on individual cases
similar to the exploration presented in this chapter. The work presented in this chapter could be developed into an evaluation framework for ambient ecologies in the future which will be beneficial for comparing results obtained in varying context, avoid reinventing the wheel and obtaining meaningful results from the user experience and usability evaluations. User experience for ambient ecology is explored in (Lino et al., 2010).

The goal with the user experience evaluation (Vermeeren et al., 2010) presented in this chapter is to get a first empirical indication of to what extent a system that operates according to some of the principles of egocentric interaction (in this case, the easy ADL ecology described in chapter 4) would meet the needs and wishes of its users. The focus is on User Experience (UX) factors (detailed out in section 5.3) rather than system efficiency or even plain usability.

Traditional usability evaluation primarily focuses on aspects like performance, ease of use and learnability. Even though such aspects are important, they are not sufficient to measure the success of an ambient ecology. An ambient ecology is not purely a system that provides certain computing services. Additional aspects like emotions, enjoyment, user affect, sensation, the meaning and value of user interactions, aesthetic experience, willingness to repeat usage, etc. play an important role, especially since human interaction within an ambient ecology typically takes place as an integral part of their everyday life and their everyday environment. In some sense, usability evaluation is included within user experience (UX) evaluation, which presents a bigger picture. According to ISO 9241-210: 2010 (clause 2.15), user experience is defined as “a person’s perceptions and responses that result from the use or anticipated use of a product, system or service”. More information about UX evaluation is given in (Hassenzahl and Tractinsky, 2006, Mäkelä and Suri, 2001). As will be shown, the evaluation of UX became challenging due to the prototypical nature of the easy ADL ecology, the complexity of the concepts explored as part of egocentric interaction, and the lack of a suitable framework for UX evaluation within ambient ecologies.

This chapter presents a user experience evaluation of performing activities supported by the easy ADL ecology (in particular the human agents’ interaction with physical and virtual objects) where user experience has been operationalized and evaluated against qualitative measurements as shown in Table 5.1. The term easy ADL ecology is used as a short expression for the term easy ADL ambient ecology.
5.2 Evaluation Setup

5.2.1 The Living Laboratory Home Environment

The living laboratory home environment used for evaluation purposes in this study can be regarded as providing a relatively realistic experience for the subjects since it is an actual home enhanced with an ambient ecology. Also, it offers the possibilities to conduct controlled studies, often difficult in a real-world scenario and often important during the developmental stages of an ambient ecology. An informal workshop was conducted with 8 subjects (students of the interaction and design program at our university) aiming to understand the needs and wishes for support in everyday activities within the easy ADL ecology. The outcome of the workshop was the development of 19 mock-up applications (refer to chapter 4) which through the aforementioned interactive and wearable mediators provide access to more than 100 virtual objects. The majority of these virtual objects were designed to facilitate the execution of three variations of everyday home scenarios that were deliberately provocative explorations of what could potentially be possible and useful in the future, rather than minor extensions to common scenarios of today.

9 smart objects with interactive mediators and 42 smart objects with sensor motes were embedded into a 54 m² apartment effectively forming a living laboratory (Kidd et al., 1999, Intille et al., 2005) complemented by three wearable mediator smart objects: two bracelets for hand gesture based virtual object manipulation and one Bluetooth wireless headset enabling speech based manipulation and audial perception (refer to Fig. 4.5 in chapter 4). The 9 smart objects acted as multimodal mediators by being equipped with LCD displays (except the wall smart object where a projector was used) of varying dimensions for visual perception of virtual objects (section 4.7.3. in chapter 4) and a pair of audio speakers for audial perception (section 4.7.4. in chapter 4). The interaction manager component (refer to chapter 4) made use of Microsoft Speech SDK 5.1 API for speech recognition of speech commands (section 4.7.2. in chapter 4) and for speech synthesis, and used a simple gesture recognition engine to discriminate among the hand gestures (section 4.7.1. in chapter 4). Apart from the smart objects, more than 300 plain objects (i.e. physical objects without any mediators) were available in the living laboratory home environment to be used by the subjects when performing everyday activities.

5.2.2 Participants

20 subjects (14 male and 6 female) not part of the research team took part in the UX evaluation with an average age of 28 years (minimum age was 18 years and the maximum age was 52 years). Even though the number of subjects used in this evaluation is limited making it difficult to generalize the results obtained, it is still a sizable total to get initial insights about the UX of the easy ADL ecology. The subjects were mostly affiliated with our university
(student or employee), belonging to different departments including medicine, engineering and social sciences. The subjects were compensated with “mouth of praise” and gifts worth around 50 SEK. Since many of the subjects were involved with research within our university, they were enthusiastic and appreciated the research efforts in building the easy ADL ecology, but at the same time they tended to look at the solutions offered from a practical point of view.

5.2.3 Scenarios within the easy ADL Ecology

Three everyday home scenarios, called weekday morning, weekday evening and weekend were created and the subjects were asked to play out part of those scenarios, performing everyday activities in a natural manner. The question of how to perform the everyday activities was left to be answered by the subjects themselves; however the subjects were asked to choose from 20 given activities for which the easy ADL ecology provides computing support. There was considerable noise during the enacted scenarios, especially while the subjects attempted to interact with the physical and the virtual aspects of the easy ADL ecology in parallel. For instance, at times subjects spoke to another subject while at the same time attempting to manipulate a virtual object using speech. At other occasions subjects used their hands to perform normal everyday physical operations and in parallel attempted to manipulate a virtual object using gesture commands.

The subjects were allowed to perform parallel activities, with the possibilities of interruption and resumption of activities. Since everyday human activities usually involve mobility, the subjects were allowed to move around freely, resulting in physical objects, mediators and virtual objects entering and leaving the subjects’ situative spaces in a dynamic fashion (refer to chapter 3). The activities were physical-virtual activities (Pederson, 2003), i.e. activities that called for actions in both the physical world and in the virtual (digital) world. For example, the activity of “preparing lunch” was performed using computing support and thus the activity ceases to be a purely traditional physical activity. Each subject performed at least five activities as part of the individual scenarios with some activities like “having dinner” or “watching a movie” also involving a second subject. Note that in such situations only one subject received the computing support offered by the easy ADL ecology, while the other subject participated in a more traditional unassisted fashion.

5.2.4 Wizard-of-Oz Method

Implementation of an ambient ecology that fully adheres to the egocentric interaction principles will remain a challenge for some time to come. For instance, accurate sensing of the situative spaces (refer to chapter 3) is hard, even if sensor technology development certainly is making it increasingly possible. Some of these technological challenges are addressed in this thesis,
like for instance activity recognition (refer to chapter 8), tracking of state changes of physical objects and virtual objects (refer to chapter 6), and tracking of the situative spaces (refer to chapter 7). For the UX evaluation, a Wizard-of-Oz method (Hudson et al., 2003, Ardito et al., 2009) was used to fill in the still missing (or not integrated) parts of the easy ADL ecology infrastructure. A Wizard-of-Oz software was implemented such that two experiment leaders (wizards) could provide information about a subject’s situational context and activity context to the personal activity-centric middleware, something which is assumed to be possible to determine automatically with reliable accuracy in the future.

The first wizard was employed in indicating the subject’s activity context by selecting one of the 20 activity contexts that were hard-coded into the software, including a “no activity” in case the subject is not performing any meaningful activity. The action context was ignored during the UX evaluation since including it would involve more work for the wizard which might lead to a reduction in their accuracy of selecting the right activity context.

A second wizard was employed in indicating the subject’s situation as viewed through the framework of the situative space model (refer to chapter 3). It was possible to hard-code into the system a fixed number of situations which included a certain number and type of objects in the various situative spaces for every situation due to: (1) the nine previously mentioned smart objects in the easy ADL ecology were stationary with fixed physical location; (2) the three wearable smart objects were always accessible to the human agent irrespective of their physical location since they were body worn; and (3) many of the plain objects in the easy ADL ecology like bed, sofa, etc. had fixed physical location. The subjects could move around freely within the home environment, and based on their location and orientation, the second wizard selected one of 32 possible physical situation options (the subjects could be in one of the 8 possible locations within the home and could be in one of the 4 possible directions namely north, east, south and west). The number of locations and directions were defined on beforehand based on the size of the home and its interior layout.

The Wizard-of-Oz method introduces errors incurred by the wizards, but the two wizards (researchers) who ran the experiments were trained for the experimentation scenarios beforehand in order to minimize the errors when the actual experiments took place. There were latency delays (a few seconds in the worst case) that could occur occasionally due to the wireless infrastructure deployed in the home environment. Such short delays should have limited consequences in affecting the quality of a human agent’s interaction within the easy ADL ecology, but the subjects were still briefed about it.
5.3 Methodology

5.3.1 Establishing the UX Factors
An important challenge in setting up the UX evaluation of the easy ADL ecology was to select the appropriate and necessary UX factors to measure (refer to Table 5.1). The underlying motivation for performing the experiment is to let the scores provided by the subjects for individual UX factors provide guidance for the further development of an appropriate ambient ecology. As a starting point for selecting the UX factors, heuristic evaluation proposed in (Nielsen and Molich, 1990) and later adapted to ambient displays by (Mankoff et al., 2003) were used. Furthermore, we also let the underlying principles of egocentric interaction (refer to section 3.2) influence the selection of UX factors. Additional UX factors were included to understand the influence of the easy ADL ecology in providing physical-virtual activity support to the human agent part of it.

5.3.2 UX is Momentary, Long-term and Anticipatory Experience
User experience is not static, but changes over time. Other factors like the human agent's earlier experiences, preconceptions, the current contextual conditions, etc. affect the UX making it subjective (Law et al., 2009). Table 5.1 presents qualitative measures that are quantized to make them comparable between users in different scenarios and useful for further analysis. Since UX is dependent on the current context and varies at different points in time, UX was evaluated both during (to a limited set of UX factors) and after the interactions within the easy ADL ecology.

Momentary UX is useful for understanding the user's affect at a particular moment in time, while long-term UX reveals the overall experience with, in this case, the easy ADL ecology. For instance, UX factors that are influenced by the subject's immediate experience like visual perception load, audial perception load, motoric load, effectiveness of the peripheral notification techniques, etc. were self-reported by the subjects (talk aloud method) to a non-intervening observer who took notes during the related events. They were afterwards discussed during the interview session to obtain the scores for the individual factors ranging from 1 to 5. UX factors that are related to the subject's remembered experience like convenience in performing physical-virtual activities, wearability and ergonomics, aesthetics and pleasing design of smart objects, agent control, uneven conditioning, etc. were self-reported after completing the scenarios by filling out a questionnaire. Imagined UX based on anticipated use and user expectation was also evaluated. For instance, a human agent's anticipated experience in various social contexts within the easy ADL ecology like being alone, being with a partner, being with friends, being with family members, etc. were self-reported and discussed during the interview session. Imagined UX of anticipated use is valuable especially since the easy ADL ecology is still in the development
phase and could be made to accommodate or adjust for the scenarios imagined by the subjects as part of the remaining implementation process.

5.3.3 Within and Without the easy ADL Ecology

Before the subjects enacted scenarios, they were briefed about the work, the living laboratory home environment, the scenarios, what to do and what not to do. The subjects trained the speech recognition engine on an average for 12 minutes and the gesture recognition engine for 10 minutes.

The subjects enacted the scenarios in two phases. During the first phase, the easy ADL ecology was not present in the sense that the living laboratory home environment was similar to an ordinary home populated with purely physical objects. The subjects could access desktop applications through a laptop (based on the WIMP interaction paradigm) for obtaining activity support. The subjects were not wearing the bracelet smart objects and the headset smart object. During the second phase, the easy ADL ecology was activated with physical and virtual objects co-existing in the living laboratory home environment. The ambient intelligence applications were providing support to the human agent based on the rules shown in Table 4.1 (refer to chapter 4) based on egocentric interaction principles. The subjects were wearing the necessary mediators for manipulating the virtual objects. The subjects took approximately 3 to 4 hours to complete the two phases, to fill out the questionnaire and to attend the interview session. The UX measures were self-reported by the subjects due to the difficulty obtaining any corresponding physiological measures. Log files for evaluating some of the technical aspects of the easy ADL ecology were generated by the smart objects and the wearable computer, especially for calculating the accuracy of the speech recognition mechanism, the gesture recognition mechanism, etc. that are described in chapter 6.

5.3.4 UX Scores and their Value

After enacting the three scenarios (both phase 1 and phase 2) mentioned earlier, each subject filled in the questionnaire form with scores between 1 and 5 for each question where the scores represent the user experience during phase 2 in comparison with their experience during phase 1. A score of 5 refers to very good (positive) experience during phase 2 in comparison to phase 1; similarly, a score of 4 refers to good experience during phase 2 in comparison to phase 1; a score of 3 refers to an experience equivalent during both phase 2 and phase 1; a score of 2 refers to bad experience during phase 2 in comparison to phase 1; and a score of 1 refers to a very bad experience during phase 2 in comparison to phase 1. Usually, if subjects provide a score of 3, then the central tendency bias might occur. In this case the subjects compare their experience with the easy ADL ecology and without it. By providing a score of 3, they agree to the nonexistence of a real difference between the two phases. Note that for experiences that are non-existing during phase 1, like for instance the match between the physical and the
virtual characteristics of the smart objects since smart objects are not a part of phase 1 scenarios, the subjects provided a score that corresponds to their experience as being part of the easy ADL ecology.

The questions were intentionally non-technical and easy to comprehend. Example questions include “How convenient was it to perform the activities? Did you need to change your location often to access information? How easy was it to interact with the virtual objects? Would you like to live with your partner in this home? Were you able to complete the activities as expected? How appropriate are the smart objects in this home setup?”, etc. Some basic technical terms like virtual objects, smart objects, etc. were explained in non-technical terms before asking them to fill in the questionnaire.

The number of subjects who answered the individual questions is presented in Table 5.1. Some questions (the momentary UX factors) were answered during the personal interview session where the subjects were free to provide their comments and clarify the notes made by the non-intervening observer while they were enacting the scenarios. This session removed some of the misconceptions that the subjects had especially in understanding the questionnaire which has resulted in a few scores that were modified by the subjects. The interview session was important in correlating the scores reported in the questionnaire to the subject’s actual experience. The quantified scores in Table 5.1 would be of significantly less value without an interview session that clarified the reasons for the scores.

5.4 Results and Discussion

Many subjects acknowledged the possible benefits of the easy ADL ecology in enhancing the everyday life of human beings in the future, especially the lifestyle of elderly and persons with disabilities. This is encouraging since one of the primary goals of this work is to explore and develop an interaction paradigm for ambient ecologies that support everyday human activities.

5.4.1 Human Physical-Virtual Activities

Providing support to human physical-virtual activities is an important feature of the easy ADL ecology. Egocentric interaction shares its part as well in looking towards an activity-centric interaction paradigm. Even though the focus has been on personal activities in this work, the long-term goal is to facilitate both personal and collaborative activities. Only 75% of the subjects reported their convenience in performing everyday activities within the easy ADL ecology. During the interview session it became clear that the significant amount of cables in the apartment, the need to wear bracelet and headset smart objects, etc., made it difficult for them to express their convenience in performing activities. However, the subjects who did answer gave an average score of 3.87, indicating their positive experience in conveniently performing activities. Many of those subjects were also aware of the prototypical nature of
the easy ADL ecology and were able to anticipate the potential convenience in performing activities in the future.

Table 5.1. User experience factors within an ambient ecology, the number of subjects responding to individual experience factors, and the average scores for individual factors between a scale of 1 and 5 is presented.

<table>
<thead>
<tr>
<th>User Experience Factors</th>
<th>Score (1 to 5)</th>
<th>No. of subjects who answered the questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Human physical-virtual activities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) Convenience in performing physical-virtual activities</td>
<td>3.87</td>
<td>15</td>
</tr>
<tr>
<td>b) Seamless integration of virtual objects within those activities</td>
<td>4.45</td>
<td>20</td>
</tr>
<tr>
<td>c) Efficiency in performing physical-virtual activities</td>
<td>4.05</td>
<td>20</td>
</tr>
<tr>
<td>2. Human situatedness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) Location change requirements to interact with virtual objects</td>
<td>3.80</td>
<td>20</td>
</tr>
<tr>
<td>b) Useful and relevant presentation of virtual objects</td>
<td>4.39</td>
<td>20</td>
</tr>
<tr>
<td>c) Human agent control in being part of the easy ADL ecology</td>
<td>4.35</td>
<td>20</td>
</tr>
<tr>
<td>d) Uneven conditioning</td>
<td>3.37</td>
<td>15</td>
</tr>
<tr>
<td>3. Human-centered load for interacting with virtual objects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) Visual perception load within the easy ADL ecology</td>
<td>4.32</td>
<td>19</td>
</tr>
<tr>
<td>b) Auditory perception load within the easy ADL ecology</td>
<td>4.79</td>
<td>19</td>
</tr>
<tr>
<td>c) Motoric load within the easy ADL ecology</td>
<td>2.89</td>
<td>18</td>
</tr>
<tr>
<td>d) Cognitive load within the easy ADL ecology</td>
<td>3.74</td>
<td>19</td>
</tr>
<tr>
<td>e) Wearability and ergonomics of the wearable unit</td>
<td>2.00</td>
<td>20</td>
</tr>
<tr>
<td>4. Human attention</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) Unobtrusiveness of the virtual objects</td>
<td>4.25</td>
<td>20</td>
</tr>
<tr>
<td>b) Effectiveness of the peripheral notification technique</td>
<td>4.20</td>
<td>20</td>
</tr>
<tr>
<td>c) Sufficient virtual objects (not too many and not too few)</td>
<td>4.18</td>
<td>17</td>
</tr>
<tr>
<td>5. Multimodal interaction with virtual objects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) Intuitive usability of the interaction techniques</td>
<td>3.65</td>
<td>20</td>
</tr>
<tr>
<td>b) Efficiency in accomplishing human initiated interaction tasks</td>
<td>3.95</td>
<td>20</td>
</tr>
<tr>
<td>c) Synchronization among multiple modalities</td>
<td>3.47</td>
<td>19</td>
</tr>
<tr>
<td>d) Easy transition to more in-depth information</td>
<td>3.90</td>
<td>20</td>
</tr>
<tr>
<td>e) Visibility of all important states</td>
<td>4.25</td>
<td>20</td>
</tr>
<tr>
<td>6. Social context</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) While being alone</td>
<td>4.95</td>
<td>20</td>
</tr>
<tr>
<td>b) While with a partner</td>
<td>3.30</td>
<td>20</td>
</tr>
<tr>
<td>c) While with friends</td>
<td>3.05</td>
<td>20</td>
</tr>
<tr>
<td>d) While with family members</td>
<td>2.90</td>
<td>20</td>
</tr>
<tr>
<td>e) While with a stranger</td>
<td>2.90</td>
<td>20</td>
</tr>
<tr>
<td>7. Smart objects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) Match between the smart objects’ physical and virtual characteristics</td>
<td>4.05</td>
<td>20</td>
</tr>
<tr>
<td>b) Aesthetic and pleasing design of the smart objects</td>
<td>4.05</td>
<td>20</td>
</tr>
</tbody>
</table>

Treating human activities as first class objects and centering human interaction with virtual objects based on their activity context has yielded a score of 4.45 indicating very good UX. Such an approach has also helped in bridging the physical and the virtual aspects of the easy ADL ecology since the human agents focus on their activities and not on whether an object is physical or virtual, as long as the object is useful in completing the current
activity successfully. Taking an activity-centric approach has also enabled the subjects to perform physical-virtual activities with good efficiency yielding a score of 4.05. Activities within an ambient ecology are likely to be physical-virtual (Pederson, 2003) and having ubiquitous access to the virtual aspects of the ecology improved the efficiency in performing those activities.

5.4.2 Human Situatedness

Human situatedness is an important principle of egocentric interaction, and the evaluation setup was designed to take the human agent’s situational context into consideration, using the situative space model to ensure that only useful and relevant virtual objects were presented to the subjects within the different situative spaces and sets, yielding a very good UX score of 4.30. A mobile subject’s relationship with environmental objects keeps changing dynamically. Even though the virtual objects are supposed to follow a moving subject, providing ubiquitous virtual object access, other contextual conditions, especially the availability of feasible mediators, affect the possibilities of ubiquitous access to virtual objects. The subjects’ score of 3.80 for location change requirement in interacting with virtual objects indicates that the experience was good, but not very good. It is due to the dual nature of smart objects: in some instances they allow for the hijacking of its mediation resources for use by virtual objects that are not associated to those smart objects, while in other instances the smart objects allow only those virtual objects associated to it. However, introducing such a trade-off allowed for maintaining a reasonably good match between the smart objects’ physical and virtual characteristics (score of 4.05).

Human mobility causes uneven conditioning (Satyanarayanan, 2001): the resources available in different human situations vary significantly. By considering a wearable approach (speech and gesture based manipulation of virtual objects using wearable headset and bracelet smart objects) to complement an augmented environment approach, a subject could interact with virtual objects even in situations without environmental mediators like on a lawn outside the house thereby masking uneven conditioning to some extent.

A human agent’s control of a situation is an important factor to consider in an ambient ecology. Ambient intelligence applications bombarding with virtual objects and services can take away a human agent’s control of their current situation resulting in human agents bowing down to the situations created by the ambient intelligence applications. The mixed-initiative approach with separate interaction management rules for human-initiated and application-initiated modes of interaction proved rather successful: the subjects reported a very good feeling of being in command (score of 4.35).
5.4.3 Human-Centered Load during Interaction

A human agent living as part of the easy ADL ecology might experience additional load on their visual perception, audial perception, motoric action, and other cognitive functions due to the inclusion of additional virtual objects and services. The easy ADL ecology is designed to make sure that such valuable resources of a human agent are not misused or over-used. Before the UX evaluation, it was assumed that augmenting a physical environment with virtual objects and services would add to the visual and audial perception load for a human agent (since the subjects need to perceive both physical and virtual objects) compared to environments with only physical objects. Surprisingly, the subjects gave a user experience score of 4.32 for visual perception experience and 4.79 for audial perception experience. This suggests that virtual objects can successfully be co-located with physical objects within the easy ADL ecology and that the unobtrusive presentation of virtual objects even may be considered as an experience enhancer.

Evaluating a human agent’s cognitive load as being part of the easy ADL ecology in comparison to being a part of an ordinary home environment is tricky since human cognitive load is influenced by a number of other factors. What is meant by cognitive load is how much the subjects had to think or keep in mind in order to interact with virtual objects within the ecology. The subjects gave a score of 3.74 for their cognitive load. During the interview session, the notes made for instance when a subject forgot the right speech command to manipulate a virtual object or when they were super efficient in performing certain activities involving a bit of thinking in manipulating physical and virtual objects (e.g. cooking) were discussed and influenced the final score of 3.74. During the interview session it also became clear that the reason they did not give a higher score was not that they felt that too much information was presented in the perception space that needed cognitive effort (which also fits the earlier mentioned score for visual and audial perception load). The reason was instead that it took some cognitive effort to recall the speech commands and the gesture commands (natural speech and gestures are not supported by the current prototype).

Motor load was given a score of 2.89, which indicates that additional efforts were required from the subjects’ perspective to perform gesture-based manipulation of virtual objects. Both speech and gesture techniques were welcomed by the subjects provided they were improved in the future.

5.4.4 Human Attention

Human attention is another important and scarce resource. This work has tried to investigate how it can be economized with a focus on peripheral attention. Human perception and action possibilities are related to human attention possibilities. The effectiveness of the proposed peripheral notification techniques both visually and audially (refer to chapter 4) is indicated by a score of 4.20, and its unobtrusiveness in making use of the
agent’s attention (score of 4.25) is promising for further investigation in this direction. In ambient ecologies, approaches that make proper use of a subject’s peripheral attention are more likely to succeed compared to those seeking the agent’s central attention all the time. Some subjects discussed about the different possibilities of presenting virtual objects so as to shift between peripheral and central attention, the most popular idea being to dynamically vary the visual objects’ size. Only 17 subjects answered if there were sufficient virtual objects (not too many and not too few) within the different situative spaces to make optimal use of the subject’s peripheral and central attention. This could be due to the rather abstract and complex nature of the question. However, those who answered were satisfied with the number of virtual objects within the situative spaces (score of 4.18).

5.4.5 Multimodal Interaction with Virtual Objects

Supporting multiple modalities for a human agent to interact within the easy ADL ecology was given high priority. It enables natural interaction in varied situations. Uneven conditioning due to the potential unavailability of appropriate mediators that could support a specific modality was masked to some extent since the virtual objects could be accessed and manipulated through several modalities depending on the contextual conditions including the availability of feasible mediators. It is important to make sure that the virtual object’s state presented in different modalities within the easy ADL ecology is synchronized. Synchronization among multiple modalities (score of 4.37) was achieved by the centralized handling of virtual objects within the easy ADL ecology. As mentioned earlier, egocentric interaction does not choose between a centralized and a decentralized approach: both approaches have their pros and cons. Other usability factors like easy transition to more in-depth information, efficiency in accomplishing human-initiated interaction tasks, intuitive usability, etc. were considered and evaluated in different contextual conditions. A score of 4.05 for efficiency in accomplishing human-initiated interaction tasks indicate that the mixed-initiative approach and the handling of the virtual objects through two types of interaction sessions (human-initiated interaction sessions and application-initiated interaction sessions) were yielding good user experience. The subjects can perform human-initiated tasks while in parallel benefitting from application-initiated virtual objects.

5.4.6 Social Context

Human beings in general are social animals and an ambient ecology might include more than one human agent. The work reported here have chosen to focus on the case of a single human agent before proceeding to investigate the case of ambient ecologies supporting multiple human agents in parallel. As part of the preparations to extend the approach to multi-agent ecologies, the subjects were asked to imagine themselves being in the ambient ecology under various social circumstances. Several issues were discussed including
the roles of individuals in a social context and the possibilities to share resources. The subjects experienced the initial attempts in addressing privacy of virtual objects within the easy ADL ecology and the scores were not just a mere speculation but a result of experiencing privacy issues and other related issues. Most subjects felt that they would have a very good experience if they were alone (score of 4.95), which is not very surprising since the three scenarios were designed for giving support to only one subject at a time (even though some scenarios were enacted by two subjects in parallel).

The subjects were doubtful if they would want to be in an ambient ecology while being with a stranger (score of 2.90): privacy, trust and many other factors come into play. Surprisingly, the subjects were also reluctant to be in such an ambient ecology while being with family members (score of 2.90), and friends (score of 3.05). The subjects expressed a need to socialize with people rather than interacting with computers even though they acknowledged that the computers within the easy ADL ecology were not like computers in the traditional sense. A deeper analysis might show that, again, the scenarios and the support provided by the ambient intelligence applications were for single human activities. Introducing scenarios where collaborative activities were performed by family members or friends could have changed the user experience scores and is an interesting future work to do. The modeling of the easy ADL ecology based on egocentric interaction does not restrict the support to be given to a single human agent since multiple situative spaces for individual human agents within the easy ADL ecology could be developed and investigated in the future. However, multi-agent ecologies raise a number of additional issues. A simple case where two human agents are concerned with their own personal activities might introduce disturbances, conflicts, resource sharing issues, etc., that a support system could detect and help to avoid or alleviate. There might actually also be support for serendipitous, ad-hoc cooperation. It is conceivable that the subjects would have had a less skeptical view with regard to the social issues if the ambient ecology had exemplified support for collaborative activities.

5.4.7 Design of Smart Objects

The design of smart objects targeted to be a part of the easy ADL ecology is a major challenge. Congruence between the smart objects’ physical and virtual characteristics is considered to be an important design principle. For instance, a bathroom smart object should not present food recipes (virtual object) in the bathroom as they are bad places for cooking, preparing a food shopping list, or almost any activity associated with the process of cooking food. The congruence principle disallows smart objects to be one-device-for-many-purposes like the typical desktop computer. An equally weighty but conflicting design principle inherited from the area of Ubiquitous Computing is that virtual objects should not be device-centric: the virtual objects in general are expected to be able to move freely and not be stuck in some particular physical object at a particular physical location. The conflict between these two design principles results in what is referred to as the dual nature of smart objects and a requirement for a proper trade-off in the design.
that allows both principles to exercise influence in parallel. The subjects have answered the UX factor of *match between the smart objects’ physical and virtual characteristics* with a score of 4.05 acknowledging the design decision to yield good experience.

Smart objects are not just devices providing functionality but should be integral parts of a human environment, so they need to be designed in an aesthetically pleasing fashion. The subjects were impressed with some of the smart objects, especially the *dining table* smart object and the *bathroom mirror* smart object, and gave an overall score of 4.05 for *aesthetic and pleasing design of the smart objects*. Two subjects asked us if some of the smart objects could be designed for their home. Inspired by (Mankoff et al., 2003) on ambient displays, the aesthetics and pleasing design was given equal importance to their functionality while designing the smart objects.

### 5.4.8 Additional Remarks by the Subjects

The interview session was important to clarify the emotions and the enjoyment that the subjects had when being part of the easy ADL ecology. 16 of the subjects enjoyed being a part of this ecology, with enthusiastic emotions and hopeful expectations on the future. The remaining 4 subjects were more cautious in expressing their feelings. The subjects’ willingness to live in the easy ADL ecology as mentioned earlier has a lot to do with their social context. *Wearability and ergonomics* were a major concern for the subjects with a score of 2.0 since wearing the currently quite ergonomically intrusive objects while sleeping, taking a shower, dressing, etc. are hard to imagine at least in the near future. It is also important to consider these issues if mild-dementia patients are expected to be a part of the easy ADL ecology since they often have the habit of removing wearable objects like wrist watch, belt, etc., when they are frustrated.

Since many of the subjects were from the department of public health at our university, there were concerns regarding the health hazards that might arise by living in the easy ADL ecology over long periods. Even though there is an awareness of the possible health hazards of radio communication, this work focuses on designing an ambient ecology and leaves it to experts in the medical field to evaluate and provide insights into the health issues.

Many subjects discussed about the loss of freedom to be the way they would like to in their home. They acknowledged that the easy ADL ecology was not pushy with too much information that forced them to behave in a certain manner, however were concerned about applications that might require a change in their behavior to fit life with the easy ADL ecology.

Egocentric interaction aims to give the human agents the ultimate control: there is no intention of introducing applications that would negatively affect human agents’ control within the easy ADL ecology. The ambient ecology components should rest silently in the background and only come in to provide support to a human agent if it believes that some support is required. Some subjects felt that the proposed ambient ecology might be detrimental to human learning and development: the opportunities for human agents to
exercise and improve their cognitive abilities. The importance of this argument is appreciated; however, it is important to be clear that the ambient intelligence applications providing support in everyday situations are expected to be configured (setting the level of automaticity, etc.) by the human agents according to their liking so that they could still exercise their minds if they wish to do so. In the case of mild-dementia patients (the long-term application domain), the intention is to leave the application configuration to a caregiver or an occupational therapist.
Chapter 6

Tracking State Changes of Physical and Virtual Objects

Keeping track of the state changes to physical and virtual objects caused by a human agent’s manipulation of those objects is an important challenge within ambient ecologies. There are several approaches to tracking the state changes to objects, however in this chapter the direction will focus on: (a) a ZigBee-based wireless sensor networking of smart objects, and (b) gesture and speech-based manipulation of virtual objects. The ZigBee-based wireless sensor network consisting of 42 smarts objects yielded an overall precision of 91.2% and a recall of 98.8% in tracking state changes to physical objects. Hand gesture recognition based on 3-axis accelerometer data obtained from the bracelet smart objects yielded a precision of 89.4% (recall of 92.2%) while speech recognition based on speech signals obtained from the Bluetooth headset smart object yielded a precision of 88.3% (recall of 90.1%) in recognizing virtual object manipulations.

6.1 Introduction

Tracking the state changes to physical and virtual objects within an ambient ecology is important contextual information for facilitating human-environment interaction and to provide contextual input to ambient intelligence applications. Tracking of state changes to physical and virtual objects caused by a human agent’s interaction with those objects is referred to as the manipulated set within the situative space model. While there are several state changes that can happen to objects, typical object state changes are caused by human manipulation either directly and indirectly. Indirect manipulation includes for instance, a microwave oven that automatically turns off after the food is heated which was initially turned on by a human agent. Tracking state changes to objects is shown to be useful in recognizing a human agent’s activities (chapter 8) and situation (chapter 7).

There are several approaches to tracking the state changes to physical objects. Scene analysis is an approach where a camera is used to capture
footage which is then sent to image recognition and analysis algorithms for recognizing objects and their state changes. Such an approach is often affected by lighting and other environmental conditions, occlusion, etc. Also, scalability becomes a problem especially in an ambient ecology comprising of several physical rooms and hundreds of everyday objects. Embedding simple state change sensors on objects provide state change information directly and are a popular approach within ubiquitous computing. In this chapter the focus is on networking smart objects embedded with simple state change sensors.

6.2 Wireless Sensor Networking of Smart Objects

This section will focus on the investigation, implementation, deployment and evaluation of a wireless sensor network of smart objects within a living laboratory easy ADL ecology.

6.2.1 What are Wireless Sensor Networks?

A wireless sensor network is made up of a large collection of sensor nodes that are usually augmented on objects in an environment where the environment could be a home, an airport, a train, a farm, etc. Wireless sensor networks are usually applied for several different purposes including monitoring, location tracking, activity recognition, security, and automation within an ambient ecology. The sensor motes sense their local environment like internal state changes to objects, conditions surrounding those objects, nearby human agents, etc. and forward such information to a destination node that receives all the information for further data processing. Since the sensor nodes are expected to operate with limited power requirements, the communication range between the sensor nodes and the destination node determines if the network should be deployed as a single hop or multihop (i.e. the sensor node transfers information to the destination node through other nearby sensor nodes through routing) sensor network.

Some research efforts for tracking the state changes to smart objects like the Mediacup (Gellersen et al., 1999), Interfacing-the-foot (Paradiso et al., 2000), AwareMirror (Fujinami et al., 2005), etc. have focused on individual smart objects instead of an ecology of smart objects. While both approaches has its merits and demerits, a wireless sensor network of smart objects can potentially provide some functionalities that individual smart objects in isolation cannot within an ambient ecology setup. For instance, recognizing the activities performed by a human agent based on manipulating objects (manipulated set) cannot be performed if smart objects exist in isolation. The information generated using the wireless sensor network of smart objects is useful for situation and activity recognition (refer to chapters 7 and 8), context-aware computing, and facilitating situated interaction (refer to chapters 4 and 5).
6.2.2 General Requirements

An informal workshop was conducted with 8 participants (either student or employee at Umeå University) in establishing the requirements for deploying a wireless sensor network connecting smart objects and a wearable computer running the personal activity-centric middleware within a living laboratory easy ADL ecology. The primary goal of this workshop was to understand what the subjects felt about placing sensors and other technological stuffs in a home environment and to get their initial reactions as a feedback. Based on the discussions, the following general requirements were imposed:

- **Non-functional (Usability, Availability and Installation):** The participants preferred a wireless sensor network that is readily available off-the-shelf at an affordable price and are easy to install. Better if the objects purchased are already equipped with sensor motes, even though such objects are yet to hit the market as mainstream products. The general opinion was that they should not be dealing with the technical issues, similar to using a telephone without knowing its backend mechanisms. While the embedded sensors should not be obtrusive in performing everyday activities, the participants felt a need to wash or clean certain objects and were curious to know what would happen. A typical example is if a coffee cup embedded with sensors could be used in a dish washer. They also preferred not to carry or wear too many devices as part of their wearable outfit (one device was informed to be the acceptable limit).

- **Performance (Functional):** While performance itself is not the main concern for the participants, they were able to understand that good performance will improve their experience of using wireless sensor networks. The participants preferred a wireless sensor network that is robust, reliable and provide adequate performance. Hence the sensing precision and recall values for the network as a whole are important evaluation aspects. Also, the participants wanted the network to have coverage over the entire home environment since their activities are distributed over the entire home environment. However, they did not require coverage beyond the home environment since many of the everyday objects do not move out of the home except for the wearable objects like hand bags, jackets, shoes, etc. The participants were not keen on charging the batteries of the smart objects often, even though they did not mind charging the wearable computer regularly as they normally do with their mobile phones.

6.2.3 System Description

The system described in this section consists of a set of smart objects present in the easy ADL ecology connected to a wearable computer running a personal activity-centric middleware described in Chapter 4. Since the easy ADL ecology contains environmental smart objects, a purely wearable approach is
not feasible, but could be used to complement the environmental wireless sensor network deployed. A wearable wireless sensor network for ubiquitous healthcare and activity monitoring is described in (Lee and Chung, 2009) and could be seen as a potential future extension to keep track of state changes to wearable objects.

There are several wireless sensor nodes with limited processing and communications capabilities including iMotes (Kling, 2003), BTNodes (Jan Beutel et al., 2004), Smart-Its (Beigl and Gellersen, 2003), Smart-Its Particles (Decker et al., 2005), etc. available to the research community both as research prototypes and as off-the-shelf products. Even though many of such motes have their specific advantages, they do not meet the general requirements for deploying them in the easy ADL ecology where state changes to physical objects like home appliances, furniture, cutlery, kitchen ware, toiletries, etc. are to be tracked. The ease of installation, usability and adequate performance in realistic setups has also been motivating in going for the deployment of a custom made wireless sensor network infrastructure taking into account the existing objects, their properties and design, and the usage scenarios in which the wireless sensor network will be evaluated.

Smart objects are embedded with stick-on nodes that sense the internal states and state changes (based on the human agent’s interaction with it) to the physical objects and transmit this information wirelessly using ZigBee communication protocol (ZigBeeAlliance, 2011) to the human agent’s wearable computer. ZigBee was preferred over Bluetooth due to its usage of low-power digital radios intended for low data rate, long battery life and secure networking applications. ZigBee supports up to 65,000 nodes on a network, introducing the possibility to include additional smart objects to be a part of the proposed system in the future. Generic communication boards are designed with easily replaceable sensor connectors to facilitate multiple sensors by only replacing the onboard sensor and microcode. Maxstream XBee 802.15.4 transceiver and Atmel ATMEGA88-20PU microcontroller are used in individual generic communication boards. The XBee transceiver operates at ISM 2.4 GHz frequency, 1mW (0 dBm) power output and allows for data rates of up to 250 Kbps. The average data rate of all the sensor nodes was 20.4 Hz (with a maximum of 100 Hz for some nodes and a minimum of 10 Hz for a majority of the nodes). The microcontroller is run at 8 MHz. The communication boards augmented within smart objects include a 2.4 GHz omni-directional antenna with \( \frac{1}{2} \) \( \lambda \) wavelength and a gain of 2.90 dBI. The generic communication boards require 3 Volts, and are powered by three 1.2V 2600mA NiMH battery in series. The receiver node connected to the human agent’s wearable computer include a Maxstream XBee 802.15.4 transceiver and a circuitry board for USB connection to a Sony vaio VGN-UX70 with 1 GHz processor and 512 MB RAM. The USB connection is considered as a serial port connection by the wearable computer.

The majority of the sensor nodes (70%) operate at a low sampling rate of 10 Hz where the nodes transmit only when there is a change in the sensor reading range defined by threshold values in the microcontroller. The internal states and state changes of the respective smart objects are calibrated by the object manager (part of the personal activity-centric middleware) based on
their unique identities. Such a double-step calibration allows for introducing additional internal states for smart objects by the object manager based on the requirements from other middleware components and the ambient intelligence applications. High sampling rate cases are currently dealt with (e.g. accelerometer (150 Hz), RFID reader for sensing a set of objects in a space surrounding the human agent, etc.) where a dedicated channel is allocation for communication between the sensor nodes and the receiver node. The sensor nodes transmit the sensed data three times (default ZigBee protocol value that could be increased, but was found to be sufficient for our purpose considering the probability of correctly receiving 113488 packets at the receiver node) to the receiver node before a time-out. According to (Tapia et al., 2006), the probability of correctly receiving packets at the receiver node increases with the number of retransmissions. This is important to consider since channel noise and collisions are issues that exist within an ambient ecology and need to be handled. However, too many retransmissions can block-up the network bandwidth, there by requiring a threshold for number of retransmissions. The microcontroller part of the sensor mote sends the data four times (experimentally found out to be the ideal number of times, but further work is required to reduce the overhead) through a USART to the XBee Transceiver and expects an ACK message in return.

The data format used for communication include Object Identity (3 bytes), Sensor Data (S1, S2...Sn), and end-of-frame (1 bit). The length of the Sensor Data depends on the smart object. (S1, S2...Sn) values are between 4 bytes (minimum length) and 17 bytes (maximum length). Sensor Type information is not included in the data format to reduce the size of the data frame. Even though the current data frame size is not an issue within the easy ADL ecology, including RFID readers (work currently being done) for modeling passively tagged smart objects within the ambient ecology might create an issue when it comes to the data frame size. The Object Identity information (refer to the ER diagram in chapter 4) is used within the personal activity-centric middleware to query a database containing information about smart objects present in the easy ADL ecology including the sensor types used for sensing their internal states. The wireless sensor network consists of a star network topology considering its simplicity (removes the need for complex routing or message passing protocols), better performance (removes the need to pass data packets through unnecessary nodes), power efficiency and isolation of smart objects that are actually not changing their internal state with smart objects that actually do so. The primary disadvantage of high dependence on the functioning of the receiver is of less importance in this context since the receiver is supposed to work 100% anyway for the personal activity-centric middleware to manage the smart objects. A star network topology demands that all the sensor nodes are within the vicinity of the receiver node. Much related work (Murty et al., 2008, Zhang et al., 2011) has deployed mesh network topologies (given the benefits of self-configuration and unlimited coverage area) which however increases the cost of the wireless sensor network and the complexity involved, making that topology type less suitable for researchers and developers that are new to networked sensors.
In the easy ADL ecology, the receiver node is worn by the mobile human agent. Since the state changes that are sensed are those created by the human agent based on their interaction with the smart objects, the sensor nodes invariably fall within the proximity of the receiver node. However, in order to cover for exceptional cases, the range within which the receiver node can receive sensor data (button on-off) with acceptable noise (< 5%) was evaluated. It turned out to be 33 meters with a single indoor wall obstruction and 19 meters with multiple (greater than two) indoor wall obstructions. Refer to Fig. 6.1a and Fig. 6.1b. For ambient ecologies that require a range greater than the ones mentioned above, repeater nodes could be easily integrated to the proposed infrastructure. 81 sensors were augmented on 42 smart objects in the living laboratory easy ADL ecology as shown in Fig. 6.2. The receiver node is designed to receive sensor data from the sensor nodes that are designed based on a combination of the basic eight sensor types described in Appendix 6.1. Additional sensor types with RS232, I2C, or SPI output can be easily included by making minor firmware modifications. Analog sensors could also be included within the sensor node, but with an
external circuit that conditions the signal to the ADC input voltage range of 3V.

The sensor data from the smart objects are received and processed by the object manager, a component part of the personal activity-centric middleware described in chapter 4. Also refer to Fig. 4.4 for further information.

Fig. 6.1b. A living laboratory easy ADL ecology with signal strength measures (Surie et al., 2008).
6.2.4 Experimental Setup

The evaluation was performed in a living laboratory easy ADL ecology prepared for ambient intelligence research at the 4th Floor MIT Huset, Umeå University. Refer to Fig. 6.1 for the plan of the easy ADL ecology. Note that this was the initial physical location of the easy ADL ecology before it was moved to an apartment in Umeå for practical reasons. Refer to chapter 7 for the floor plan of the new easy ADL ecology. Chapter 8 focuses on activity recognition with the easy ADL ecology where kitchen activities were mainly chosen for evaluation considering the number and variety of activities possible in a kitchen environment. This section will follow a similar track by focusing on kitchen activities; however other activities beyond the kitchen domain are also included.

Four subjects (not part of the research team) were recruited for the experiments. They were compensated with food and mouth-of-praise. The experiments were performed individually by individual subjects for a week's duration. In addition to the sensing system's wearable components (a wearable computer running a personal activity-centric middleware and a ZigBee receiver node), the subjects were also given a wearable camera connected to a mobile digital video recorder (DVR) for obtaining the ground truth (state changes to smart objects) as shown in Fig. 6.3. The sensor readings (not used for evaluation), if an event performed by the subject was

Fig. 6.2. Sensor nodes augmented on smart objects in the living laboratory easy ADL ecology (Surie et al., 2008), also presented in chapter 4 for better readability.
not captured on the wearable mini camera were discarded. A similar wearable mini camera and DVR set-up was used for ethnographical studies on “how people perform everyday activities in a home environment?” with acceptable reliability (Bhatt, 2007). Experience sampling method (Intille et al., 2003) is a commonly used method to obtain the ground truth where the subject is supposed to manually enter the events produced by them while performing activities onto a PDA (usually) running the experience sampling method software. However such an approach was shown to be inaccurate for obtaining the ground truth from the subjects due to wrong event selection; not using the experience sampling method software for short activities like toileting; delay between sensor firing and experience sampling; etc. (Tapia et al., 2004). Indirect observation of sensor activations for activity recognition, situation monitoring, etc. results in a situation where the sensing errors are comfortably missed due to the unavailability of ground truth (Tapia et al., 2004).

**Fig. 6.3.** A subject wearing a wearable computer and a ZigBee receiver during the evaluation of the wireless sensor network within the easy ADL ecology. Note that the camera and the video recorder were used for obtaining the ground truth and is not part of the sensing system (Surie et al., 2008).

The wireless sensor network was evaluated using scenarios consisting of everyday activities within the easy ADL ecology. The data collected will especially be used for activity recognition in the future with the knowledge of precision and recall values of the wireless sensor network. The subjects’
wearable camera was used to collect the ground truth for the set of 10 activities presented in Table 6.2. Time-based synchronization was used to map the sensor firings with the ground truth. The subjects were not restricted on how to perform the activities, however were briefed on how to use the smart objects in the easy ADL ecology. For the two sensitive activities of toilet routine and changing clothes, the subjects were part of the evaluation process in identifying the ground truth where the wearable camera was not used, and instead self-made notes were used. The set of 10 activities were performed between a minimum of 7 times and a maximum of 20 times during a week. The subjects were briefly interviewed after the scenarios for clarifications. Human made errors in interacting with smart objects, the flimsy nature of some of the sensor augmentations, obtrusiveness caused by wires and sensors, etc. are examples of issues discussed.

6.2.5 Evaluation

6.2.5.1 Tx-Rx Range and Signal Strength Measures

The transmission-reception (Tx-Rx) range is usually evaluated in an outdoor environment with line of sight and free of obstructions. However, such evaluations are of less importance for the easy ADL ecology populated with smart objects, plain objects and a human agent in an indoor environment with indoor walls that can affect the transmission-reception range. By proposing a mobile receiver that is part of a human agent’s wearable outfit and is more often than not within the range of the sensor nodes activated based on the human agent’s interaction with the concerned smart object, the challenge of maintaining a good transmission-reception range was established. The transmission reception range was evaluated in an indoor environment yielding 33 meters with a single wall obstruction and 19 meters with multiple wall obstruction as mentioned earlier. The signal strength measure described in Fig. 6.1 was measured in the home environment with a certain amount of background noise, created when the smart objects like the fridge, microwave oven, regular oven, vacuum cleaner, etc. are turned on. Since the easy ADL ecology was evaluated within the university premises, there were other wireless networks outside the context of the wireless sensor network described in this section that created background noise. In Fig. 6.1, one of the subjects was asked to be at the location near the dining hall marked by Rx. The signal strength from the various sensor nodes were evaluated by a push button (on-off event) recorded 5 times at 8 different locations in the home environment. Table 6.1 shows that 97.5% of the times, the signal strength values at the 8 different locations are acceptable (>10 dB). In both the “toilet – toilet corridor” case and the “bedroom – bedroom corridor” case, the state of the door (closed or open) seems to be an important factor to consider. Similarly, the line-of-sight cases (living room, dining hall and kitchen) have performed better than cases having wall obstructions.
6.2.5.2 Sensing Precision and Recall

The accuracy in sensing the smart objects’ internal state changes based on the human agent’s interaction with those objects is an important evaluation factor. The wireless sensor network was evaluated within scenarios where the subjects were performing a set of everyday activities within the easy ADL ecology. Such an approach was decided for due to the following evaluation criteria: 1) the sensing system should be evaluated as a whole (ecology of smart objects) instead of sum of the individual smart objects in isolation; 2) the sensing system should be evaluated in a realistic setup where the subjects are performing everyday activities by interacting with smart objects subjected to various kinds of noise; and 3) the data collected from the sensing system will be used for developing and evaluating activity recognition algorithms with the additional information about the accuracy of the sensing system known. In chapter 8, multiple Hidden Markov Models (HMMs) were used in parallel to accommodate individual information channels like selected set, intra manipulation and extra manipulation for recognizing human activities and actions. It should also be noted that the activity recognition systems to be described in chapter 8 were evaluated in an immersive virtual reality simulated easy ADL ecology. Precision and recall values are defined as follows:

\[
\text{Precision} = \frac{\text{the number of correctly recognized items}}{\text{the number of recognized items}} = \frac{\text{True Positives}}{\text{True Positives + False Positives}}
\]

### Table 6.1. Evaluation of signal strengths at 8 different locations within the easy ADL ecology with the subject located near the dining hall as shown in Fig. 6.1 (Surie et al., 2008).

<table>
<thead>
<tr>
<th>Location Number</th>
<th>Location Name</th>
<th>Distribution of Signal Strength (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Best (&gt;30 dB)</td>
</tr>
<tr>
<td>1</td>
<td>Living room</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>Dining hall</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>Kitchen (door open)</td>
<td>80</td>
</tr>
<tr>
<td>4</td>
<td>Bedroom corridor</td>
<td>60</td>
</tr>
<tr>
<td>5</td>
<td>Bedroom (door closed)</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>Toilet (door closed)</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>Toilet corridor</td>
<td>60</td>
</tr>
<tr>
<td>8</td>
<td>Office (door open)</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 6.2 shows that the sensing system has an overall precision value of 91.2% and an overall recall value of 98.8%. The results are promising considering the amount of background noise present in the environment (for wireless communication) and the fact that sensing the internal state changes of some of the smart objects was tricky. For instance, the ambient light present in the human agent’s environment affected the decision of determining if the dust bin (uses light sensor) was full or empty. Hence, there was a need for performing ambient light noise cancellation for cases involving the light sensor. The location, number and type of sensors augmented on smart objects are important factors to address for obtaining high performance measures. Sensor fusion is an interesting approach to reduce the sensing errors, which will be performed in the future.

6.2.5.3 Installation Time and Initial Usability

The subjects wanted off-the-shelf products. However, smart objects with wireless sensor networking capabilities are still restricted to research community and their commercialization as products would be a step in the future. The smart objects developed were presented as prototypical off-the-shelf products with a need to install and fine-tune them to make them adaptive to the deployed environment. The installation time in this context refers to the time taken in adjusting the location of the sensor placement on objects, fine-tuning the sensor thresholds like for instance with the light sensor to know if the microwave oven is on or off (this should be done in the firmware but it is as simple as changing some numbers), learning to charge the batteries, etc. that are important for the subject to be able to manage in the absence of the wireless sensor network developer. The installation time was measured for 2 subjects separately. One subject took 65 min while the other subject took 45 min which needs to be improved in the future. The wearable computer running a personal activity-centric middleware and connected to a ZigBee receiver weighs 0.632 Kg with dimensions of 15*10*4 Cm³. All other sensor nodes are instrumented in the environment instead of including them in the human agent’s wearable outfit. Note that the wearable camera and DVR were used only for the evaluation purpose and is not part of the tracking system described in this section 6.2.

6.2.5.4 Discussion

RFID technology could be used to complement the wireless sensor network in an ambient ecology with multiple human agents. Since the objects in an ambient ecology will be shared by human agents, the state changes caused to an object by a specific human agent is important information in establishing the manipulated set for that particular human agent. Even though a wireless sensor network can establish state changes to objects, the identity of the
human agent causing such a state change is difficult to establish which can be tracked using RFID technology. The combination of wireless sensor networking with RFID technology was thought about in tracking the intra manipulation information channel used for activity recognition in chapter 8.

Table 6.2. Precision (in percentage) and recall (in percentage) values in sensing the state changes to physical objects within the easy ADL ecology while performing a set of 10 everyday activities (Surie et al., 2008).

<table>
<thead>
<tr>
<th>Activity Number</th>
<th>Activity Name</th>
<th>Sensing Precision (in %)</th>
<th>Sensing Recall (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Drinking coffee</td>
<td>84.1</td>
<td>98.8</td>
</tr>
<tr>
<td>2</td>
<td>Baking cake, bread, etc.</td>
<td>100.0</td>
<td>98.8</td>
</tr>
<tr>
<td>3</td>
<td>Doing the dishes</td>
<td>90.0</td>
<td>100.0</td>
</tr>
<tr>
<td>4</td>
<td>Repairing the coffee machine</td>
<td>74.7</td>
<td>88.0</td>
</tr>
<tr>
<td>5</td>
<td>Changing clothes</td>
<td>91.0</td>
<td>99.0</td>
</tr>
<tr>
<td>6</td>
<td>Heating-up the frozen food</td>
<td>93.6</td>
<td>100.0</td>
</tr>
<tr>
<td>7</td>
<td>Toilet routine</td>
<td>99.0</td>
<td>99.0</td>
</tr>
<tr>
<td>8</td>
<td>Preparing dinner</td>
<td>87.6</td>
<td>99.0</td>
</tr>
<tr>
<td>9</td>
<td>Setting-up the table</td>
<td>100.0</td>
<td>95.7</td>
</tr>
<tr>
<td>10</td>
<td>Having dinner</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td>Global percentages</td>
<td>91.2</td>
<td>98.8</td>
</tr>
</tbody>
</table>

6.3 Tracking State Changes to Virtual Objects

Tracking state changes to virtual objects is another important challenge within the easy ADL ecology. While virtual object state changes are easy and straight forward to keep track of since the virtual objects are part of the virtual world created by computers, the actual problem is to keep track of their state changes based on natural human manipulations. In our work, virtual object manipulation based on speech and hand gestures are kept track of. It is important to accurately track a human agent’s speech and gesture based interaction with virtual objects. Inaccurate tracking will affect the overall functioning of the easy ADL ecology (designed based on egocentric interaction) and affect human agent’s user experience. For more information about the gesture and speech based interaction techniques, refer to chapter 4.

The 20 subjects who participated in the user experience evaluation described in chapter 5 took part in this evaluation as well. On average, the subjects took 10 min for getting used to gesture commands and 12 min for
training the speech recognizer and getting used to commanding using speech. Handwritten notes were made by a non-intervening observer, complemented by log files generated by the individual smart objects for obtaining the precision and recall values of the gesture and speech commands with the virtual objects. Note that subject-performed errors were discarded while calculating the precision and recall values. True positive, false positive and false negative scores were recorded while the subject enacted the scenarios. The subjects enacted the scenarios first without the support of the easy ADL ecology and then with its support to compare the experiences. Since the living laboratory home environment is a different environment to the subjects’ home environment, the subjects were exposed to all the objects in the environment including the smart objects and the virtual objects (part of the ambient intelligence applications). It gave the subjects an idea of what to expect from the environment and how to perform physical-virtual activities in this environment. The subjects performed human-initiated interaction tasks like “obtaining the current weather information” while performing the activity of dressing; “finding information about cultural events in the city” while discussing culture with a friend; etc.

Table 6.3. Precision and recall values for gesture and speech based manipulation of virtual objects within the easy ADL ecology where TP, FP, FN, P, and R refer to True Positive, False Positive, False Negative, Precision and Recall respectively.

<table>
<thead>
<tr>
<th>Virtual Object manipulation modality</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>P (in %)</th>
<th>R (in %)</th>
<th>Human agent performed errors (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hand gestures</td>
<td>732</td>
<td>87</td>
<td>62</td>
<td>89.4</td>
<td>92.2</td>
<td>6.5</td>
</tr>
<tr>
<td>Speech</td>
<td>776</td>
<td>103</td>
<td>85</td>
<td>88.3</td>
<td>90.1</td>
<td>5.3</td>
</tr>
</tbody>
</table>

Table 6.3 shows quantitative evaluation results for gesture- and speech-based selection and manipulation of virtual objects. Initially, during a pilot study, much lower precision and recall values were obtained for the speech modality, due to background noise. In the current version of the speech recognizer, “confirmation commands” were introduced. A locking mechanism was also introduced so that only virtual objects within the perception space and/or the action space can be interacted with using speech. Even though it restricts the possibilities of bringing a virtual object from the world space to the perception space or the action space instantly (takes one or two additional speech commands), it has resulted in higher precision value for speech commands. If the basic interaction techniques using speech and gestures are not accurate then the idea of exploring an interaction paradigm, in this case egocentric interaction which is more conceptual would be a harder task.
Appendix 6.1. Summary of the sensor types used in the easy ADL ecology with their performance parameters (Surie et al., 2008).

<table>
<thead>
<tr>
<th>Sensor type</th>
<th>Measures</th>
<th>Sensor(s)</th>
<th>Range</th>
<th>Resolution</th>
<th>Power consumption</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature</td>
<td>Temperature</td>
<td>Thermometer (Dallas DS 18S20)</td>
<td>-55°C to +125°C</td>
<td>±0.5°C</td>
<td>Active mode 1 to 1.5mA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Sleep mode 750 to 100mA</td>
</tr>
<tr>
<td>Ambient light</td>
<td>Ambient light intensity</td>
<td>Omni directional light sensor (TACOS TSL2565B)</td>
<td>0 to 255</td>
<td>137 mV/W (10μm²) at 650nm</td>
<td>Active mode 11 to 1.7mA</td>
</tr>
<tr>
<td>Pressure pad</td>
<td>Where pressure is applied within an area</td>
<td>Mesh of pressure sensors (Omron microswitch SS-5) spread across a surface</td>
<td>Depends on the size of the pressure pad. A 50cm by 10cm pressure pad was used</td>
<td>4cm²</td>
<td>Active mode 1mA</td>
</tr>
<tr>
<td>Touch / Press button</td>
<td>Button touched / pressed</td>
<td>Push button</td>
<td>Binary 0 or 1</td>
<td></td>
<td>Active mode 1mA to 3mA</td>
</tr>
<tr>
<td>Touch / Press pad</td>
<td>Buttons touched / pressed</td>
<td>Mesh of push buttons spread across a surface</td>
<td>Depends on the size (no. Of push buttons included) of the touch / press pad. 35cm by 15cm, and 20cm by 10cm pads were used</td>
<td>Resolution of the touch / press button</td>
<td>Depends on the power consumption of a push button and the number of push buttons in the mesh</td>
</tr>
<tr>
<td>Appliance feedback light</td>
<td>Feedback light intensity</td>
<td>Light sensor (TACOS TSL2565R) made unidirectional</td>
<td>0 to 255</td>
<td>137 mV/W (10μm²) at 650nm</td>
<td>Active mode 11 to 1.7mA</td>
</tr>
<tr>
<td>Open / Close</td>
<td>Light intensity</td>
<td>Mesh of light sensors (TACOS TSL2565R) spread across surfaces</td>
<td>Depends on the size (no. Of light sensors included) of the object. 15. 75 decimeter³ to 270 decimeter³ volumes were worked with</td>
<td></td>
<td>Depends on the power consumption of a light sensor and the number of light sensors in the mesh</td>
</tr>
<tr>
<td>Containment sensor</td>
<td>Light intensity</td>
<td>Mesh of light sensors (TACOS TSL2565R) spread across surfaces</td>
<td>Depends on the size (no. Of light sensors included) of the object. 6.6 decimeter² to 44 decimeter³ volumes were worked with</td>
<td></td>
<td>Depends on the power consumption of a light sensor and the number of light sensors in the mesh</td>
</tr>
</tbody>
</table>
Chapter 7

Situative Space Tracking across Multiple Modalities

Situatedness is an important principle of egocentric interaction. An ambient ecology modeled using egocentric interaction concepts are naturally expected to address the challenge of tracking a human agent’s situation that spans across the physical-virtual domain. This chapter presents the research efforts taken in identifying the different modalities through which objects are perceived and acted upon, and the different approaches taken in tracking and framing the objects (both physical and virtual) within the situative spaces across multiple modalities. A WLAN signal strength-based object proximity tracking with reference to a human agent complemented by other parameters stemming from the capabilities of the individual objects is presented as the primary approach for situative space tracking. A preliminary “proof-of-concept” evaluation of this tracking system by 2 subjects in a living laboratory easy ADL ecology yielded a global precision of 83.4% and recall of 88.6%.

7.1 Introduction

Situatedness is an important principle of egocentric interaction and the easy ADL ecology modeled using egocentric interaction is expected to keep track of a human agent’s situation and use it for facilitating interaction and providing activity support. Ambient ecologies are expected to be context-aware and human situation is a particular case of context. Keeping track of a human agent’s situation provides an understanding of their symbolic location which is important contextual information within an ambient ecology. Knowing a human agent’s situation enables facilitating situated interaction with ambient intelligence applications (refer to chapters 4 and 5). Another application is to use the information generated by the situative space tracking system for activity and action recognition as described in chapter 8. Context-aware computing (Schmidt, 2002) is another important field where information generated by the situative space tracking system could be used to derive higher level context and share such context information with interested entities.
Two important properties of the situative space model described in chapter 3 are:

- Includes both physical and virtual objects, thereby providing a picture of the physical-virtual situation (Pederson, 2003) of a human agent at a particular instance in time.
- Covers the multiple modalities through which objects could be perceived and acted upon.

The situative space tracking approaches to be described in this chapter should handle the above mentioned two properties. Keeping track of both physical and virtual objects is something useful within mixed-reality research (Milgram and Kishino, 1994), especially for integrating the physical and the virtual aspects of a physical-virtual environment and for analyzing a human agent’s physical-virtual situation. However, related research efforts seem to focus on modeling and tracking a human agent’s situation either within a physical environment or within a virtual environment, instead of considering their physical-virtual environment (Pederson, 2003). Also, existing research approaches for situation modeling and tracking are driven by the technological advancements instead of basing it on conceptual models centered on a human agent. In this chapter, the situative space tracking systems to be described use the situative space model as a conceptual base. While the importance of technological advancements is appreciated, it is not considered as the starting point.

### 7.1.1 SSM across Multiple Modalities

Fig. 7.1 presents a visualization of the situative space model (SSM) (refer to chapter 3) that spans across multiple modalities including visual, audial, touch and gesture. Human agents perceive and act not based on a single modality but through multiple modalities that interact with each other to create a unified representation of perception and action. This introduces a need to not only let the situative space model span across multiple modalities, but also to operationalize mechanisms for multimodal fusion. The situative space tracking system I proposed in section 7.3 will focus on tracking the perception space (visual and audial modalities), recognizable set (visual modality), and action space (touch modality). The sensing of selected and manipulated sets using simple state change sensors was described in chapter 6. Note that in Fig. 7.1, PS (v) refers to the perception space along the visual domain while PS (a) refers to the perception space along the audio domain.

### 7.1.2 Establishing the Situative Space Boundaries

To operationalize the situative space model, mechanisms to establish the boundaries for the various situative spaces and sets are needed. The situative space model used for tracking the situative spaces and establishing their boundaries are a simplification with potential improvement possibilities. In
addition to the type of errors that may occur due to imperfect modeling of the human agent's perception and action possibilities, errors of the situative space model itself, there will be errors that occur due to the imperfections of the technological infrastructure used for tracking the situative spaces. At this time the accuracy of available, feasible technologies is inadequate, but development in this area is rapid.

Fig. 7.1. A Situative Space Model spanning across multiple modalities like visual, audial, touch, and gesture (Surie et al., 2010a).

### 7.2 Approaches to Situative Space Tracking

There are several approaches to tracking the situative spaces that includes both physical and virtual objects, and spans across multiple modalities. Based on an initial investigation it was evident that several tracking systems are required in parallel to complement each other in tracking the situative spaces that spans along several domains. A list of six different situative space tracking systems, in various stages of implementation and evaluation is presented. They are categorized according to their role in maintaining the situative spaces up to date and what kind of modality they operate in (Table 7.2). Note that in Table 7.2, the (+) is used to represent that the tracking systems operate (or should operate) in parallel complementing each other to
track the specific situative space while the (/) is used to represent that either one of the tracking systems could be used to track the specific situative space.

1. **ProxyTrack** (Surie et al., 2010a) is the situative space tracking system I to be described in section 7.3. It is the primary approach taken to track the situative spaces. It is based on WLAN signal-strength measures at the different smart objects in the easy ADL ecology with reference to a wearable access point worn as a pendant by a human agent. The signal strength measures are used to determine the proximity of the objects with reference to the human agent, and this information is combined with the individual object characteristics in framing the objects within the respective situative spaces.

2. **OrientTrack** (under development) is an infrared signal-strength-based orientation tracking system for determining the set of physical objects and mediators that are visually perceivable by a human agent based on their orientation relative to the agent’s head. An array of infrared LEDs is embedded in physical objects and mediators. An infrared camera (a modified Nintendo Wii remote) is worn by the human agent on the forehead, connected to the personal wearable computer via Bluetooth. The LEDs on each object are turned on and off to create unique patterns for each object. Since infrared technology requires free line-of-sight for communication, objects within the agent’s field of vision are easily identified. orientTrack is intended to complement ProxyTrack in determining the perception space through sensor fusion.

3. **IdTrack** (under development) is a tracking system based on passive RFID tags aimed at identifying physical objects and mediators selected by the human agent (idTrackHand) as well as those in the immediate vicinity of the agent’s body (idTrackChest). Two Skyetek’s M1-mini wearable RFID readers (5.8cm to 8.5cm range) are worn on the agent’s wrist, while one Skyetek M7 RFID reader (1m to 2m range) with a specially designed antenna for long-range passive RFID-tagged object tracking is worn on the human agent’s chest. The RFID readers are connected to the agent’s wearable computer running the personal activity-centric middleware.

4. **StateTrack** (Surie et al., 2008) is a ZigBee-based wireless sensor networking system (described in chapter 5). Physical objects and mediators in an environment are equipped with simple state-change sensors like touch sensors, temperature sensors, light sensors, etc., that determine state changes of objects caused by a human agents’ manipulation. Sensor motes were designed using Maxstream XBee 802.15.4 transceiver and Atmel ATMEGA88-20PU micro-controllers. As mentioned earlier, the system uses a star networking topology.
with sensor motes transmitting state/state-change information to the wearable computer.

5. **SpeechTrack** is a speech recognition system using Microsoft Speech SDK 5.1 API for speech recognition and a BTH-8 Bluetooth microphone worn by the human agent or microphones embedded in the environment. This system filters out speech commands from the real-world sound stream, allowing for selection and manipulation of virtual objects. For further information and user experience evaluation results, refer to chapters 4 and 5.

6. **GestureTrack** is a gesture recognition system based on Phidgets 1059 3-axis accelerometers worn on the hands of the human agent. Similar to speechTrack, the system filters out explicit gesture commands from the flow of everyday physical actions. gestureTrack currently works with a predefined set of gestures but it is intended to include end-user-defined gestures in the future. For further information and user experience evaluation results, refer to chapters 4 and 5.

Table 7.2. Different situative space tracking systems that complement each other in tracking the physical and virtual objects within the different situative spaces and sets. Note that the tracking systems sense objects spanning across multiple modalities including visual, audial, touch and gesture (Pederson et al., 2011).

<table>
<thead>
<tr>
<th>Situative space / set</th>
<th>Visual</th>
<th>Audial</th>
<th>Touch</th>
<th>Gesture</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PO</td>
<td>VO</td>
<td>PO</td>
<td>PO</td>
</tr>
<tr>
<td>Perceived space</td>
<td>Proxy Track</td>
<td>Orient Track</td>
<td>Proxy Track</td>
<td>Proxy Track</td>
</tr>
<tr>
<td>Recognizable set</td>
<td>Proxy Track</td>
<td>Orient Track</td>
<td>Proxy Track</td>
<td>Proxy Track</td>
</tr>
<tr>
<td>Examinalble set</td>
<td>Proxy Track</td>
<td>Orient Track</td>
<td>Proxy Track</td>
<td>Proxy Track</td>
</tr>
<tr>
<td>Action space</td>
<td>Proxy Track</td>
<td>Proxy Track</td>
<td>Proxy Track</td>
<td>Proxy Track</td>
</tr>
<tr>
<td>Selected set</td>
<td>Proxy Track</td>
<td>State Track</td>
<td>Proxy Track</td>
<td>State Track</td>
</tr>
<tr>
<td>Manipulated set</td>
<td>Proxy Track</td>
<td>State Track</td>
<td>Proxy Track</td>
<td>State Track</td>
</tr>
</tbody>
</table>
7.2.1 Discussion

7.2.1.1 Perception Space

This work on determining the perception space has so far focused on the visual modality (considering the dominance of visual modality in human perception), with the exception of audial modality for perceiving the presence and state of virtual objects. Hence there are empty cells along the touch modality. Refer to the perception area of Table 7.2.

Since ProxyTrack senses the presence of both physical objects and mediators, it provides information about the perceptibility of physical as well as virtual objects in the visual domain. To determine the perceptibility of virtual objects more information is needed. A mediator’s rotation (critical for 2D displays) and proximity relative to the human agent’s chest is encoded in the WLAN signal strength as measured by the mediator and filtered using a signal-strength threshold determined empirically for each visual display. A virtual object’s presence in perception space is then determined by its visual perception threshold value, its size on the display, and the size of the display itself – all known to the related smart object. The mediator-orientation information provided by ProxyTrack is to be complemented with more accurate information provided by OrientTrack.

ProxyTrack also plays an important role for determining the perception space in the audio modality. The current implementation considers the relative proximity of loudspeakers (determined by ProxyTrack), the sound volume they can generate, and the contextual noise level, to determine whether a virtual object represented in the audio modality is within the perception space. The audially recognizable and examinable sets could be formed by applying empirically determined proximity thresholds.

The perception space, recognizable set and the examinable set supports visual modality and are closely associated to unconscious human eye movements (Bulling and Gellersen, 2010). However, conscious eye movements could be used to select and manipulate virtual objects facilitating eye-based interaction as described in (Bulling and Gellersen, 2010). Hence visual modality could potentially be used for tracking the action space, selected set and manipulated set by tracking conscious human eye movements.

Cognition-aware interfaces make use of a human agent’s eye movements as valuable contextual information about the environment and associating it to their cognition (Bulling et al., 2011) which is similar to the way the concepts of recognizable set and examinable set emerged within the situative spaces.

7.2.1.2 Action Space

For the action space, the choice of gesture and speech as input modalities in the interaction infrastructure gives the SpeechTrack and GestureTrack systems important roles in enabling the manipulation of virtual objects. The speech action space outlines the range within which objects can be
Situative Space Tracking across Multiple Modalities

The presence of a suitable mediator (microphone) together with the state of the interactive system (whether it is in speech input mode) can be used to determine the presence of virtual objects in the speech action space. The gestural action space and its subsets are by and large defined analogously to the speech action space. Refer to chapter 4 for speech and gesture based interaction techniques for manipulating virtual objects.

The touch action space outlines the range within which physical and virtual objects can be manipulated using touch. ProxyTrack is currently used to approximate the tracking of objects in this space, soon to be complemented by IdTrackChest. For the selected set, StateTrack is currently used but is limited to objects capable of communicating state changes to the interactive system running on the wearable computer (refer to chapter 6). In order to include also less advanced physical objects, IdTrackHand will be introduced in the near future, allowing for any object tagged with a passive RFID tag to be accurately included in the selected set once grabbed by the human agent.

Determining whether the selected object has actually been manipulated and has changed its internal state is however still requires StateTrack. The process of inclusion of virtual objects in the touch action space follows the same lines as for physical objects, only that they depend on selection and manipulation of a mediator first, associating it with a virtual object. This work is only in its initial phases of exploring the structure of this interaction design space, which includes the use of any tagged physical object as mediator or physical “token” in the manner of Tangible User Interfaces (Shaer et al., 2004), provided by the combination of GestureTrack and IdTrackHand.

7.3 Situative Space Tracking System I

Traditionally, human agents occupy physical environments populated with physical objects. However, within an ambient ecology human agents are situated in a physical-virtual environment surrounded by co-existing physical and virtual objects creating physical-virtual situations (Pederson, 2003), also referred to as mixed-reality situations. Since smart objects are a part of an ambient ecology, they provide access to virtual objects through the mediators embedded in them. This allows physical and virtual objects to be co-present in time and space from a human agent’s perspective. Co-located physical objects (for instance, a wash basin below the bathroom mirror smart object) might not be augmented with mediators, but their spatial co-location with other smart objects allows them to also (indirectly) be a part of the ambient ecology (simply hard-coded into the system). Such an approach eliminates the need to augment all physical objects with mediators, allowing them to exist as plain objects but still is a part of the ambient ecology.

This section will describe the WLAN signal-strength-based situative space tracking system I for capturing the perception space, recognizable set and action space. The situative space tracking system I comprises of a set of smart objects and a wearable unit running a personal activity-centric middleware. The smart objects are embedded with ASUS WL-167g WLAN adapters and thin-client boards for wirelessly communicating with an off-the
shelf wireless access point (WRT54GL) that is worn by a human agent with a customized directional antenna augmented on it. The antenna is shielded with aluminum to make it directional, i.e. works mainly in front of a human agent’s chest with poor data communication with other devices located to the back of a human agent. Indoor location tracking approaches requiring high level accuracy for tracking an object’s physical location usually use three access points. However, to track the situative spaces where the object’s physical location in absolute terms is not required can be handled using a single access point. Refer to Table 7.1 for precision and recall values obtained using this single access point.

The wearable computer was simulated with a notebook computer. Java Wireless Research API developed by Johan Kristiansson from Luleå University of Technology is used by the individual smart objects to calculate the WLAN signal strength measures that indicate their proximity (and orientation to some extent) with reference to the human agent’s body. Since the situative space model itself is based on human agent’s proximity to surrounding objects (refer to the proximity principle), a proximity tracking infrastructure seems appropriate.

Signal strength-based approaches for proximity measurement is close at hand in ambient ecology setups that usually already possess wireless networks thereby removing the need for additional infrastructure. The off-the-shelf radios embedded in the smart objects serve as an object-tracking infrastructure as well as for data communication. Since physical indoor walls play an important role in determining the boundaries of the situative spaces (an approximation), all indoor walls were embedded with a layer of aluminum foil to dampen the signal strength measures of smart objects behind the walls that could not be physically perceived or acted upon at the moment by a human agent.

The situative space tracking system I do not provide physical proximity measures in terms of distance between a human agent and the smart objects. Distance measures are not required for the current purpose. Instead, the situative space tracking system I measures physical proximity in terms of signal strength measures which are directly used for tracking objects within the situative spaces. Note that to calculate distance measures, additional information about indoor walls, signal propagation characteristics, etc. is required.

### 7.3.1 System Architecture

Fig. 7.2 shows the architecture for the situative space tracking system I. The proposed architecture is primarily based on the principles of egocentric interaction. A feed forward model is used for simplicity. The architecture comprises of five components. Received Signal Strength Indicator, Situative Space Filter and Multimodal Mixer are components implemented in the individual smart objects. Situative Space Aggregator and Situative Space Stabilizer are part of the personal activity-centric middleware (the situation monitor component within it). Refer to Fig. 4.7 in chapter 4. Low-level and smart object dependent calculations are made by the smart objects, thereby
reducing the wireless bandwidth requirements in exchanging data with the middleware. Low-level calculations are not computationally heavy, and since the smart objects already possess computational power to for instance display virtual visual and audio objects, it justifies the system design. A star network topology is used for wireless communication because of its simplicity and by considering the network range requirements within the easy ADL ecology which is to cover the entire area using a single hop. In the evaluation section, it will be clear that using a star network was sufficient for the purpose intended.

7.3.1.1 Received Signal Strength Indicator

Detecting proximity based on received signal-strength indicator (RSSI) is highly sensitive to scattering, reflection, signal attenuation and measurement noise (Hashemi, 1993). Moreover, RSSI values fluctuate over time and the stability of signal-strength measures varies among different heterogeneous smart objects. Signal patterns of the different smart objects were collected during a training phase creating what is referred to as smart object fingerprints. The proximity of the smart object with reference to the human agent is measured with low accuracy due to a lot of noise. The smart objects are classified as belonging to a specific space or set based on the relatively inaccurate proximity measure. Smart object fingerprints varies among the different smart objects in the easy ADL ecology depending on their location and other external factors like nearby objects, human movement patterns near to that smart object, etc. During the tracking phase, the patterns were statistically used to discard signal strength measures that were inappropriate and inconsistent over time. Information about factors that influence signal propagation within the easy ADL ecology like indoor walls, influence of other objects, etc. were implicitly encoded within the smart object fingerprints. $SO_{\{1, 2, 3 \ldots n\}}$ represents the set of smart objects that are a part of the easy ADL ecology. The smart objects measure signal strengths at 5 Hz, and apply appropriate statistical methods to remove inconsistent signals before sending the average RSSI measure to the situative space filter every 2 seconds.

7.3.1.2 Situative Space Filter

As described in chapter 4, a smart object is a physical object which is augmented with mediators that provide access to virtual objects. The signal strength thresholds for physical object (and co-located physical objects), mediators and virtual objects are used to filter them within the situative spaces. Note that $SO_i$ for instance, might be perceived by a human agent through multiple modalities, introducing a need to filter the objects across multiple modalities. Depending on the modality, other characteristics specific to individual objects affect their position within the situative spaces. One such characteristic is the object size along the visual modality. Bigger objects could be perceived from a larger distance compared to smaller objects. Physical objects do not vary their size (if we ignore exceptional cases) with ease, in contrast to virtual objects. Virtual object characteristics like object loudness (in the case of audial modality) or object size could be varied dynamically and
the cost involved in such changes are often negligible. This makes it challenging for the situative space tracking system I to position the virtual objects within the situative spaces. At present, virtual objects that do not change their characteristics dynamically are experimented with. However, future versions should handle dynamic changes to virtual object characteristics.

**Fig. 7.2.** Situative space tracking system I architecture with WLAN signal strength used for tracking perception space, recognizable set and action space within the easy ADL ecology (Surie et al., 2010a).

Refer to Fig. 7.2. The situative space filter SSF$_I$ considers the RSSI$_I$ value across the visual, audial and touch domains separately. Within the individual domains, the RSSI$_I$ value is first fed through a Physical Object Discriminator to position the physical object (and co-located physical objects) within the situative spaces. Thresholds like V$_{PS}$ (Visual Perception Space), V$_{RS}$ (Visual Recognizable Set), V$_{ES}$ (Visual Examinable Set), A$_{PS}$ (Audial Perception Space), A$_{RS}$ (Audial Recognizable Set), A$_{ES}$ (Audial Examinable Set), and T$_{AS}$ (Touch Action Space) are used. The RSSI$_I$ value is then fed through a Virtual Object Discriminator to position the virtual object(s) within the situative spaces. This is done in two steps where the mediator(s) within the SO$_I$ is first positioned within the situative spaces based on thresholds like M$_{VPS}$ (Mediated Visual Perception Space), M$_{VRS}$ (Mediated Visual Recognizable Set), M$_{VES}$ (Mediated Visual Examinable Set), M$_{APS}$ (Mediated Audial Perception Space), M$_{ARS}$ (Mediated Audial Recognizable Set), M$_{AES}$ (Mediated Audial Examinable Set) and M$_{TAS}$ (Mediated Touch Action Space).
Then the virtual objects characteristics like object size (for visual virtual objects), object loudness (for audial virtual objects), etc. are mapped to the mediators’ positions to finally position the virtual objects within the situative spaces. It should be noted that since proximity was measured with low accuracy and the object orientation was ignored, there were negligible difference in the thresholds for the physical objects and the mediators. So for the evaluation to be described in section 7.2.2, the same threshold values for the physical objects and their respective mediators were used (refer to Fig. 7.4 for the signal strength thresholds of the individual smart objects).

In the future, additional contextual information like background light, background noise, etc. could be used to further filter objects within the situative spaces. The situative space filter SSF outputs the following sets that contain information about the physical objects, virtual objects and mediators within their respective situative spaces and modality.

7.3.1.3 Multimodal Mixer

The Multimodal Mixer fuses the objects spread across different modalities within individual situative spaces. Such a fusion is motivated by a human agent’s ability to perceive or act on an object through multiple modalities. Physical objects to some extent are easier to handle computationally compared to virtual objects that might possess multiple manifestations. Individual objects have their natural characteristics that enable a human agent to perceive and act on those objects in specific modalities. For instance, an alarm clock is usually perceived through the audial modality (at least while performing its intended purpose of sending out an alarm). The multimodal mixer should effectively combine the natural bias of individual objects to specific modalities, and at the same time use redundancies as a way of improving the robustness (i.e. eliminate noise and sensing errors). Issues of correlation between modalities that enhances or affects the perception and action possibilities of individual objects should be considered by the multimodal mixer. For instance, an object that is both visually and aurally perceivable could be given higher probability of being in the fused perception space compared to an object that is only aurally perceivable. (Welch and Warren, 1980) propose the hypothesis that processes involving multiple modalities follow modality appropriateness, where visual modality dominates over audio modality or touch modality for spatial tasks, while audio modality dominates for temporal tasks. Multimodal Mixer MM for instance fuses \( PS_v \) and \( PS_a \) to obtain \( PS_t \) that represents the set of physical objects, virtual objects, and mediators that are part of smart object SO within the perception space. It should be noted that the evaluation described in section 7.2.2 to evaluate the perception space (visual modality), perception space (audial modality), recognizable set (visual modality) and action space (touch modality) of the situative space tracking system I does not make use of the multimodal mixer in calculating the precision and recall values for the
individual situative spaces across specific modalities. The multimodal mixer is still part of the architecture since it is important to fuse the modalities and calculate the accuracy of the situative spaces as a whole spanning across multiple modalities, left for the moment as future work.

7.3.1.4 Situative Space Aggregator

The Situative Space Aggregator is responsible for receiving situative space information from different smart objects that are in the close proximity to a human agent. $PS_1$ to $PS_n$ for instance are aggregated to $PS$ representing the global perception space. During the aggregation process, relationships among objects (both physical and virtual) could be used to improve accuracy and to compensate for the lack of mediators in some smart objects. For instance, the co-location of objects could be used, i.e. certain set of objects are co-located within the easy ADL ecology. Co-location of objects could be considered along both spatial and temporal domains. Magic Touch (Pederson, 2001) is an object location tracking system based on RFID and ultrasound technologies where the concept of containment is used to co-locate objects within a container with the container itself. For instance, if papers are within a tracked folder and the folder is moved to a new location, then the papers are also assumed to be in the new location. Several approaches exist to establish the co-location relationship among objects including the passive RFID-based approach, and a hard-coding approach based on manual input of spatial relationships between physical objects. The latter approach, also used in our design suffers from obsolete relationship problems in a dynamic environment where the objects are not stationary, but removes the need for augmenting mediators within plain physical objects.

7.3.1.5 Situative Space Stabilizer

Human agents regularly move their bodies or body parts in such a way that it changes their proximity and orientation towards objects within the easy ADL ecology. This creates scenarios where objects enter and leave the situative spaces dynamically. Sometimes even too fast for a human agent to perceive or act on them. Human agents do visual stabilization of the objects that enter and leave their perception space using fixational eye movements (Martinez-Conde et al., 2004). Their visual system makes use of the fact that physical objects do not flicker in and out within physical environments. Hence, the role of the situative space stabilizer is to stabilize the objects (impose a temporal delay before objects are regarded as present) within the situative spaces by checking their presence or absence over a period of time which is more appropriate for the visual modality. In the touch modality, once again the human agent’s proximity to an object that enables touching and manipulating it over a period of time is considered before putting the object in the action space.

The hypothesis is that such an approach ensures that an object for instance, within the perception space is actually perceivable by a human
agent. The stabilizer also removes noise due to signal fluctuations that are common within an ambient ecology, thereby making sure that only stable objects are positioned within the individual situative spaces. The situative space stabilizer introduces time domain to the situative space model. The stabilization time is to some extent dependent on the application, and also the objects involved. For instance, a complex architectural diagram might impose larger stabilization time compared to a simple image of a flower to visually perceive. sPS for instance, represents information content within the perception space that is stabilized over a period $S$. The stabilization time for audial modality needs further research since the presentation of audial virtual objects takes place over a period of time.

7.3.2 Evaluation I

The easy ADL ecology populated with 9 smart objects (only 9 smart objects were used since the remaining smart objects either work based on ZigBee protocol forming a wireless sensor network or are wearable smart objects that communicate using Bluetooth protocol as described in chapters 4 and 6) representing a set of physical objects (and collocated physical objects), mediators and virtual objects was used for evaluating the situative space tracking system I. The easy ADL ecology in its present form supports only a single human agent within it. Refer to Fig. 7.3. Multi-human agent ecologies pose additional challenges like situative space sharing, resolving conflicts within the situative spaces, privacy issues, etc. which are beyond the scope of this thesis.

Two subjects took part in the experiments. The subjects were students in our department with some experience of using computers. They were given a brief introduction to our work, and the concepts of smart objects, perception, action, etc. were described. Information that could potentially lead to biased results was not discussed. The subjects provided the ground truth using the talk aloud method, while hand-written notes were made by a non-interfering passive observer (one of the researcher). To understand the decisions made by the situative space tracking system I, time-stamped log files were obtained from the individual smart objects and the personal activity-centric middleware.

Windows Media Player 9.0 was used as the virtual environment containing a variety of virtual objects like text objects, audio objects, video objects, image objects, etc. that are distributed across multiple modalities. The experimentation took place in two stages: a training phase and a recognition phase. During the training phase, the situative space thresholds were established, while during the recognition phase the situative space tracking accuracy was established.
Wireless Coverage

The Transmission-Reception (Tx-Rx) range is important to make sure that the proposed tracking system has 100% coverage within the easy ADL ecology. A Tx-Rx range of 72 meters (with -65 dB signal strength as a threshold to maintain network quality) was obtained in the living laboratory home backyard (outdoor environment). However, such an evaluation is of lesser importance compared to their performance indoor. A Tx-Rx range of 48.27 meters was obtained as an average (among 10 readings) with a single indoor

Fig. 7.3. A living laboratory easy ADL ecology populated with smart objects (only smart objects used by the situative space tracking system I are shown), a personal activity-centric middleware, ambient intelligence applications and a human agent in the middle of it all (Surie et al., 2010a).
wall obstruction within the 54 m² easy ADL ecology, while the network still showed acceptable quality for multiple indoor wall obstructions.

7.3.2.2 Signal-Strength-based Thresholds

The subjects positioned themselves in 32 different locations in the easy ADL ecology, and spent 6 to 8 minutes in individual locations and were constantly changing their proximity and orientation to establish the spatial representation of the RSSI values with reference to the surrounding smart objects. The subjects provided the ground truth for the individual situative spaces along different modalities in which a particular physical object and/or virtual object (accessed through a mediator) was present every 5 to 10 seconds. Note that the virtual object characteristics like size, loudness, etc. was kept fixed. Even though the study is a preliminary one with significant amount of labor work required, and the thresholds are valid for this particular ambient ecology setup (especially the number of smart objects used and their locations) with further studies required, it is still useful in calculating the thresholds for the situative spaces for individual smart objects as a proof-of-concept. Future efforts of establishing the signal strength thresholds could be benefited by developing a model of the environment. Refer to Fig. 7.4. The variation among the individual subjects is not considered in calculating the thresholds, but would be useful to consider in the future, especially if this study will involve more number of subjects with varied background making the situative space tracking system personalized to individuals.

7.3.2.3 Precision and Recall Values

The subjects performed everyday activities in the easy ADL ecology as part of different scenarios at home. Such everyday activities resulted in constant changes within their situative spaces. Ground truth about the situative spaces (objects belonging to individual spaces and sets with their modality information) was obtained every 10 to 20 seconds and was compared with the log files generated. The subjects preferred to provide the ground truth while being stationary compared to while being mobile. Since the intensity of interaction between a human agent and their environment is low while being mobile (something that was made note of during the evaluation), the influence of the absence of proper ground truth in such situations could be considered negligible. Even though the evaluation performed is limited in the number of subjects and the duration spent by each subject, which was about 60 minutes each, the results obtained so far (refer to Table 7.1) are promising to continue further research in this direction. The visual modality has yielded good results both for the perception space (precision of 99.80%) and for the recognizable set (precision of 99.78%). Audial modality has a precision of only 63.62% due to background noise and the fact that the subjects were not confident enough in distinguishing audio information within the situative spaces compared to visual information. It should be noted that since both the subjects did not find a real difference between the recognizable set and the
examinable set in the visual modality, and between the perception space and the recognizable set in the audial modality, the examinable set (visual modality) and the recognizable set (audial modality) were discarded from the evaluation.

**Fig. 7.4.** Situative space thresholds for the individual smart objects in the living laboratory easy ADL ecology represented as signal strength measures in dB (Surie et al., 2010a).

Action space has a precision of 100%, but a recall value of 55.03% shows that the system was not able to make a decision 44.97% of the time which is too high a value and could be reduced by considering complementary tracking technologies like passive RFID technology. Thus, the attempt to validate a theoretically grounded situative space model (refer to chapter 3) in practice using WLAN signal strength-based tracking is presented within the easy ADL ecology context. It should be noted that the global precision and recall values in Table 7.1 is based on averaging out the results for individual situative spaces and sets without using the multimodal mixer component described in section 7.2.1.

The accurate tracking of objects (and mediators) and framing them within the individual situative spaces will probably remain a challenge for years to come. While this chapter does not claim to have solved the challenge of situative space tracking across multiple modalities, it reports one concrete step in this direction with the situative space tracking system I (or ProxyTrack).
Table 7.1. Precision and recall values (in %) of the situative space tracking system I for the different situative spaces and set. Final row represents the global values including global precision and recall values considering the situative space tracking system I as a whole (Surie et al., 2010a).

<table>
<thead>
<tr>
<th>Situative space / set</th>
<th>True positive</th>
<th>False positive</th>
<th>False negative</th>
<th>Tracking precision (in %)</th>
<th>Tracking recall (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perception space (visual)</td>
<td>743</td>
<td>3</td>
<td>16</td>
<td>99.6</td>
<td>97.9</td>
</tr>
<tr>
<td>Perception space (audial)</td>
<td>717</td>
<td>410</td>
<td>86</td>
<td>63.6</td>
<td>89.3</td>
</tr>
<tr>
<td>Recognizable set (visual)</td>
<td>450</td>
<td>1</td>
<td>31</td>
<td>99.8</td>
<td>93.6</td>
</tr>
<tr>
<td>Action space (touch)</td>
<td>164</td>
<td>0</td>
<td>134</td>
<td>100.0</td>
<td>55.0</td>
</tr>
<tr>
<td><strong>Global values</strong></td>
<td><strong>2074</strong></td>
<td><strong>414</strong></td>
<td><strong>267</strong></td>
<td><strong>83.4</strong></td>
<td><strong>88.6</strong></td>
</tr>
</tbody>
</table>
Chapter 8

Recognizing Human Activities and Actions

Activity-centeredness is an important principle of egocentric interaction. An ambient ecology modeled using egocentric interaction concepts are naturally expected to address the challenge of modeling and recognizing human activities that spans across the physical-virtual domain. This chapter investigates the feasibility of the different information channels generated using the situative space model for activity recognition. Perception space, action space and selected set were used by the activity recognition system I, while selected set and manipulated set (both intra-manipulation caused by internal state changes to objects and extra-manipulation caused by external state changes to objects) were used by the activity recognition system II. An immersive virtual-reality simulated easy ADL ecology was used for the initial, yet concrete evaluation yielding promising results. Discussions on activity recognition and the potential areas of improvement in the future are presented.

8.1 Introduction

Activity-centeredness is an important principle of egocentric interaction and the easy ADL ecology modeled using egocentric interaction is expected to recognize human activities and use it as first-class information for providing activity support and facilitating interaction. Also, keeping track of a human agent’s situation and activities implicitly using sensor and recognition algorithms simplify the understanding of a human agent’s intention and their immediately needs useful within an ambient intelligence context. Over the years, activity recognition has also become an important challenge to address within ambient intelligence research.

Activity recognition has several applications within an ambient ecology context. Typical applications include healthcare and elderly care at home where an activity recognition system is used for monitoring health status, providing emergency alarms, preventive healthcare, and supporting the successful performance of Activities of Daily Living (ADL). The easy ADL ecology is developed within the context of the easyADL project (Backman, 2006) for providing ADL support to patients suffering from mild-dementia.
disease. Other applications of activity recognition include providing training to newly employed staffs in a manufacturing plant, providing assistance for specialized activities like surgery in an hospital, supporting athletes during their training sessions, supporting passengers in a modern airport, improving shopping experience in a supermarket and many more.

Developing an activity recognition system is a challenging task due to:
1. The number and variety of activities performed by human agents.
2. The variations in the way different human agents perform similar activities.
3. The variations in the way a specific human agent performs an activity over time.
4. The possibilities of performing several activities in parallel.
5. The possibilities of constant shifting among several activities resulting in dynamic suspension and resumption of activities.
6. The possibilities of performing an activity in several locations and often while in roaming.

While activity recognition in itself is a hard challenge, recognizing human activities in environments occupied by several human agents increases the complexity.

8.1.1 Approaches for Activity Recognition

There are several approaches to recognizing human activities.

1. **Body and limb movements**: activity recognition based on a human agent’s body and limb movements using accelerometers is described in (Bao and Intille, 2004, Lee and Mase, 2002, Kern et al., 2003). This approach is restricted to recognizing a subset of everyday human activities involving body and limb movements like walking, running, bicycling, brushing teeth, taking an elevator, etc. Complex activities that involve object manipulations common in an ambient ecology setup are hard to recognize using this approach. Typically, body-worn sensors are used for this approach and are a popular approach in the wearable computing community.

2. **Audio and sound**: activity recognition based on audio is described in (Chen et al., 2005). Such an approach is restricted to a subset of sound-related activities. Typically a microphone is used for sound recordings which is usually considered obtrusive and can affect privacy. Since everyday environments can be noisy, noise cancellation mechanisms are required to improve recognition accuracy.

3. **Camera and image processing**: recognizing activities using single but complex sensors like a camera is described in (Haritaoglu et al., 2000). Such an approach is severely hampered by the complexity in extracting valuable features, even though human beings use the sense of vision effectively. Even though this approach to activity
recognition is one of the earliest, capturing images and video using a camera can be obtrusive and affect user privacy.

4. **Object manipulation:** recognizing activities based on object manipulation by a human agent (here the reference is to the selected set according to the situative space model) is becoming an important approach (Philipose et al., 2004, Patterson et al., 2005) within ubiquitous computing community. Human agents perform activities by grabbing and realizing objects which is used as an information channel for recognizing activities. This approach enables recognizing a wide range of complex everyday activities performed human agents in an ambient ecology setup. Typically, everyday objects are usually augmented with passive RFID tags, while wrist-worn RFID bracelets are used to keep track of the objects that are grabbed and released by a human agent. This approach suffers from the lack of sharp events that might occur based on object manipulation and could be useful for activity recognition.

5. **State changes to objects:** recognizing activities based on state changes to objects are another important approach within the ubiquitous computing community. This approach enables recognizing even a wider range of activities in comparison to the object manipulation approach. Tracking state changes to objects yield sharper events and enable activity recognition during the initial phase when an activity has just begun. Also, determining the end of an activity, suspension and resumption of an activity, etc. are simplified using this approach. Typically, everyday objects are augmented with sensor motes that capture state changes and wirelessly transmit this information to an activity recognition engine. The main limitation of this approach is that all the objects that are useful for activity recognition are to be augmented with sensor motes affecting the simplicity of everyday objects.

While many activity recognition systems are heavily driven by currently available sensor technology, the approach take in this thesis is to build activity recognition systems starting out from a perspective more centered on how humans literally perceive the world and based on the weight that current cognitive science gives to this source of information as an activity-driving factor. In other words, the activity recognition systems are based on the principles of egocentric interaction with a particular focus on the situative space model. We believe that our approach could offer a conceptual design platform for activity recognition robust enough to survive and handle generations of changes in the field of sensor technology.

Activity theory (Nardi, 1996a) introduces a 3-level hierarchy of activity, action and operation in describing an activity. An activity takes place in several situations, where each situation is comprised of a set of actions under certain conditions like location, time, etc. An action is a conscious goal-directed process performed by a human agent to fulfill an objective and is comprised of a set of operations. Operations are unconscious processes that depend on the structure of the action and the environment in which it takes
place. The main goal of this chapter is to develop a general approach for modeling and recognizing human behavior with finer granularity: not only at the activity level, but also at the action and operation levels.

The feasibility of the different information channels generated by the situative space model for recognizing human activities and actions is investigated in this chapter.

- Recognizing activities based only on information about objects being grasped or released is a promising approach, but there is no previous work (to the knowledge of the author) that has used this information alone in recognizing human actions which are at a lower abstraction level compared to human activities. However, an extension of such an approach (activity recognition system I) using the situative space model has shown promising results (refer to section 8.2).

- As mentioned earlier, recognizing activities based only on objects grasped or released has also had difficulties in recognizing the activities during their initial phase, when the activity has just begun. This introduces long delays between the actual starting of an activity and the moment the activity recognition system makes a guess about the human agent’s current activity, a problem that has been addressed to the extent possible by activity recognition system II described in section 8.3.

- Also, the approach based on object manipulation has had difficulties in recognizing the end of an action or an activity due to the unavailability of sharper events that could be generated using information channels for intra manipulation and extra manipulation, which will be discussed in section 8.3.

8.2 Activity Recognition based on Perception Space, Action Space and Selected Set

8.2.1 An Egocentric Perspective of Everyday Objects

As mentioned in chapter 3, the term ‘egocentric’ has been chosen to signal that it is the body and mind of a specific human agent that serve as center of reference in modeling their interaction with physical and virtual objects within an ambient ecology. Human agents situate themselves closer to the objects relevant for their current activity. Such explicit situatedness gives an indication of the human agent’s intent (the needs and wants of the human agent to satisfy some goal). In this chapter it is shown that by capturing the changes within a human agent’s perception space and action space (part of the situative space model described in chapter 3) it is possible to indirectly capture their intentions represented in the activities performed by them. The situative space model serves as a conceptual tool for determining what situative aspects of human activities need to be captured and in what detail.

The prototype to be described in this section (8.2) uses the presence of objects within a human agent’s perception space and action space along with the human agent’s selection of objects by grabbing and releasing. The
prototype to be described in section 8.3 uses the states and state changes to objects caused by the human agent’s manipulation of those objects. A comparison of the two activity recognition systems show that the first recognition system deals with information channels that are at a coarse granularity level in describing a human agent’s interaction with situated objects while the second recognition system deals with information channels that are at a fine granularity level in describing a human agent’s interaction with situative objects.

The current version of the situative space model (the one described in chapter 3) does not consider other human agents within the situative spaces and sets, while in the future the model will be extended to include other human agents. Such an extension will be useful for modeling and recognizing collaborative and group activities, and also for facilitating activity support for individual activities in a multi-agent ambient ecology where resources in the environment need to be shared.

### 8.2.2 Operationalization of the Situative Space Model

The situative space model is operationalized as follows for providing input information channels to the activity recognition system I.

- **The world space** contains the set of all objects part of the ambient ecology that is known to the personal activity-centric middleware described in chapter 4. In the immersive Virtual Reality (VR) setting that is used for the evaluation purposes, it is the set of all objects included in the VR model simulating an ambient ecology.

- **The perception space** (PS) is the set of objects within a cone in front of the human agent’s eyes with this cone following the head movements as shown in Fig. 8.1 (left). The height of the cone is limited by the walls in an indoor environment and visual occlusion further affects the number of objects within this space. Note that visual modality alone was used in operationalizing the perception space used for activity recognition.

- **The action space** (AS) is the set of objects within a hemisphere in front of the human agent’s chest (Fig. 8.1, right). Such a shape is motivated by the fact that humans have two hands and the assumption that they manipulate objects within reach of their hands. The hemisphere follows the human agent’s chest movements and its radius is equal to the maximum distance between the chest and a hand. Note that touch modality alone was considered in operationalizing the action space used for activity recognition.

- **Selected set** (SS). When objects are manipulated by a human agent, objects are either selected or de-selected resulting in two types of events: `objectID_grasped` and `objectID_released`. Although the human agent can manipulate objects with both hands, we do at the moment not make any distinction between right and left hand manipulations. A similar approach is described in (Philipose et al., 2004, Patterson et al., 2005). The selected set represents the operations performed by a human agent during the accomplishment of an action (and activity) by selecting or de-selecting objects.
Immersive virtual reality is used as a test bed in order to speed up the design (and re-design) process, and to compensate for the limitations with the currently available sensing technologies for tracking the situative spaces. A VR model developed using the Colosseum3D real-time physics platform (Backman, 2005) is used to simulate wearable sensors and sensors on everyday objects within the easy ADL ecology.

Fig. 8.2 shows a snapshot of the VR environment. Perception space and action space are captured at 1 Hz, while the selected set is captured when a grasp or release event occurs. Refer to Fig. 8.3 for the system architecture. The evaluation (refer to section 8.2.3) was conducted using 70 object types. Object types include mobile object types like fork, knife, plate, etc., and stationary object types like microwave oven, sink, stove, etc. Only the type of object is considered in recognizing activities, not the identity (e.g. fork_1 and fork_2 are both considered as fork type). There are many objects that overlap for several activities. For instance, fork, knife, plate, etc. are used for several activities like preparing breakfast, preparing the table, having breakfast, and doing the dishes. This makes the classification problem harder compared to taking an approach where the recognition system is strongly characterized by one or two objects that are unique to the activity. For some activities like for instance preparing_rice, there is a unique object, in this case rice_bag. But this does not simplify the classification problem for the following reasons: 1) the activity recognition system I not only recognizes the human agent’s current activity, but also the human agent’s current action, and the rice_bag is not manipulated by a human agent in all the actions within this activity, but only in a few of them; 2) the rice_bag manipulation might be noise inadvertently created by the human agent while performing another activity; and 3) the recognition system should recognize the activity and action before they are actually completed to provide activity-centered assistance to the
human agent. Hence the system cannot wait until the unique object is manipulated to recognize the activity and action.

**Fig. 8.2.** Perception space, action space, selected set and manipulated set captured in an immersive VR simulated easy ADL ecology.

**Fig. 8.3.** Activity Recognition System I architecture where the information channels A (perception space), B (action space) and C (selected set) are combined (Surie et al., 2007b).
8.2.3.1 Feature Extraction

Perception space and action space both consist of sets of objects that need to be quantified. The quantification scheme builds $S_A$ and $S_B$ as shown in Fig. 8.3, where the vectors represent the list of distinct object types with their corresponding number of occurrences. A log function is applied on $S_A$ and $S_B$ to give more importance to the type of objects present within those spaces compared to the number of their occurrences. One limitation of such a quantification scheme may be its scalability to a large number of object types, since the dimension of the quantification vector depends on the total number of object types. Having experimented with 70 types of objects, the results obtained were good.

8.2.3.2 Clustering

$F_A$ and $F_B$ are fed into two distinct clustering algorithms to retain the features provided within their respective information channels. The cluster center for perception space ($O_A$) and action space ($O_B$) is obtained which is used for further classification as shown in Fig. 8.3. The growing neural gas (GNG) clustering algorithm (Fritzke, 1995) was preferred to classical clustering techniques like K-Means or Kohonen SOM because it maintains a steady learning rate and also creates new clusters with additional training data. The end-user is not expected to fine-tune or parameterize the system for new situations added. However, in the current implementation, the GNG algorithm is used as a classical clustering algorithm with a fixed number of clusters. The above mentioned features will be included in the future.

8.2.3.3 Classification

The probabilistic generative framework of a hidden-markov model (HMM) (Rabiner, 1989) is used because of its clear Bayesian semantics, its ability to handle time-varying signals and the availability of efficient algorithms for state and parameter estimation. HMMs reduce the system’s configuration space into a number of finite discrete states together with the probabilities for transition between the states. One limitation of HMMs is that the model structure has to be user-defined, which includes the number of states and the connections between the states. The model structure cannot be determined by standard learning methods. This should not pose a major problem since the activities recognized are user-defined. The human agent provides ground truth for both activities and actions. Each activity is modeled using a separate HMM with the number of states corresponding to the number of actions within that activity. Similarly, the transitions between states correspond to the transitions between different actions of an activity. HMMs have shown good results in many activity recognition systems including (Chen et al., 2005, Lukowicz et al., 2004, Lester et al., 2005, Patterson et al., 2005). The activity recognition system I uses three information channels (refer to Fig. 8.3). Each information channel produces a sequence of observations that are fed into ten HMMs (one for each activity). For each information channel, the outputs from
the ten HMMs are used to build an activity probability vector \((A_A, A_B \text{ and } A_C)\) containing the probabilities for each possible activity. The element of the activity probability vector with the highest value gives the human agent’s current activity and its most probable state gives the human agent’s current action.

### 8.2.3.4 Combining the Information Channels

The three information channels are combined using activity contribution factors \((W_A, W_B \text{ and } W_C)\) and action contribution factors \((w_A, w_B \text{ and } w_C)\). \(W_A, W_B \text{ and } W_C\) consist of the recognition precision values for each activity while \(w_A, w_B \text{ and } w_C\) consist of the recognition precision values for each action. These factors are automatically generated from the training data. The human agent’s current activity is first determined by computing \(R\), the weighted sum of all the three information channels:

\[
R = W_A \times A_A + W_B \times A_B + W_C \times A_C
\]

where * represents element-by-element multiplication. The element of \(R\) with the highest value gives the human agent’s current activity \(A\). Once the activity is known, the human agent’s current action is determined by calculating \(r\) using the following formula:

\[
r = w_A \times pS_A(A) + w_B \times pS_B(A) + w_C \times pS_C(A)
\]

where \(pS_A(A), pS_B(A) \text{ and } pS_C(A)\) are the vectors of state probabilities of the HMM representing the human agent’s current activity \(A\) for perception space, action space and selected set respectively. \(w_A, w_B \text{ and } w_C\) are equal to zero if their respective channel is not supportive to the activity determined previously. The element of \(r\) with the highest value gives the human agent’s current action. The three information channels were not combined before classification, in order to evaluate the individual information channels independently; and also to allow for additional channels without affecting the overall infrastructure of the activity recognition system.

### 8.2.4 Evaluation I

The quantitative evaluation was performed having 4 subjects in the immersive virtual reality easy ADL ecology. 10 activities were included as shown in Table 8.1.

The activities were performed 20 times each as part of various scenarios. A scenario comprises of a few related activities performed in some sequence. Example scenarios include breakfast scenario, lunch scenario, and free-time scenario. Some activities were common for several scenarios like the activity of doing the dishes which is common to both the lunch scenario and the breakfast scenario. The subjects were allowed to perform the activities in their own style (often in many different ways). When a subject begins performing their activity, each object is in the location where it was last placed in the
subject’s previous activities. This makes the experiments realistic compared to having a fixed initial location for each object. Cases when the subjects dropped an object on the floor or took the wrong object were also included in the dataset. A real chair was used for the subjects to perform the activities of having breakfast and having lunch. Subjects’ body postures and locomotion within the virtual reality environment were realistic. For instance, the subjects were not allowed to pass through a table, even though it is possible in a virtual reality environment.

**Table 8.1. List of actions and activities selected based on the AMPS framework (AMPS, 2011).**

<table>
<thead>
<tr>
<th>Activity Number</th>
<th>Activity Name</th>
<th>Actions within individual Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Preparing rice</td>
<td>Get the rice bag. Pour rice into the cooker. Pour water into the cooker. Add salt. Put back the rice bag</td>
</tr>
<tr>
<td>2</td>
<td>Preparing fried vegetables</td>
<td>Get some vegetables. Cut those vegetables. Fry those vegetables. Add spices. Place the chopper in the sink</td>
</tr>
<tr>
<td>3</td>
<td>Preparing cake</td>
<td>Get the baking plate. Add some eggs. Add some milk. Add some sugar. Add some cake powder. Place the baking plate in the oven</td>
</tr>
<tr>
<td>4</td>
<td>Preparing coffee</td>
<td>Take some coffee powder. Pour water into the coffee machine. Get some cups. Pour some coffee into the cups</td>
</tr>
<tr>
<td>5</td>
<td>Preparing breakfast</td>
<td>Toast some bread slices. Boil some eggs. Prepare some juice. Prepare the cereals</td>
</tr>
<tr>
<td>6</td>
<td>Doing the dishes</td>
<td>Clean the dishes. Dry the dishes on the rack. Wash the hands</td>
</tr>
<tr>
<td>7</td>
<td>Having lunch</td>
<td>Have the main meal. Have the dessert. Drink coffee. Place the used dishes in the sink</td>
</tr>
<tr>
<td>8</td>
<td>Having breakfast</td>
<td>Have the main meal. Place the used dishes in the sink</td>
</tr>
<tr>
<td>9</td>
<td>Preparing the table</td>
<td>Place the table mats. Get some cutlery. Get some plates. Get some glasses. Get the food. Place some napkins</td>
</tr>
<tr>
<td>10</td>
<td>Cleaning the kitchen</td>
<td>Clean the table. Clean the stove. Clean the rack. Clean the floor</td>
</tr>
</tbody>
</table>

According to (Ward et al., 2011), evaluation of activity recognition systems should be objective and unambiguous, independent of parameters like the number of activities, number of actions within each activity, number of objects in an environment that enable performing activities, etc. Also, the evaluation should be more than a simple binary quantitative evaluation where results like “the precision of system A is better than system B” adapted from the field of pattern recognition is valuable yet insufficient. The evaluation should highlight the strengths and the weaknesses of the activity recognition system, and its applicability to the specific domain of interest (Ward et al., 2011).

The number of clusters and the observation sequence length were empirically determined for individual information channels based on the recognition accuracy. Activity recognition system I was trained and tested on recorded data, using various combinations of these parameters. The optimal
parameters are not sharply defined, allowing a variation of around 10% without altering the results significantly. Refer to Table 8.2.

**Table 8.2.** Optimal parameters for perception space, action space and selected set (Surie et al., 2007b).

<table>
<thead>
<tr>
<th>Optimal parameters</th>
<th>Perception space</th>
<th>Action space</th>
<th>Selected set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of clusters</td>
<td>55</td>
<td>70</td>
<td>No clusters</td>
</tr>
<tr>
<td>Sequence length</td>
<td>15</td>
<td>15</td>
<td>7</td>
</tr>
</tbody>
</table>

### 8.2.4.1 Precision, Recall and Confusion Matrix

“Leave-One-Out Cross-Validation” (LOOCV) scheme was used to obtain the precision and recall figures (see Table 8.3). Precision and recall are defined as follows:

\[
\text{Precision} = \frac{\text{the number of correctly recognized items}}{\text{the number of recognized items}} = \frac{\text{True Positives}}{\text{True Positives + False Positives}}
\]

\[
\text{Recall} = \frac{\text{the number of correctly recognized items}}{\text{the number of true items}} = \frac{\text{True Positives}}{\text{True Positives + False Negatives}}
\]

On evaluating the information channels independently, perception space shows the best results, both at the activity level and the action level. A deeper analysis of the clusters formed within perception space shows that it contains more than simple situative information. For the same perception space (e.g. “Around the stove”), several clusters are formed for different activities (e.g. “Stove + objects for cooking”, “Stove + ingredients for preparing cake”). The clusters actually contain information about both the situation (object locations) and the activity, considering the objects brought by the human agent while performing the activity. So the clusters formed in perception space are not only snapshots of patterns in the environment, they also represent the dynamic changes made by the human agent when they perform the activities. This also explains why the optimum number of clusters is high for both perception space and action space. Action space works in a similar manner, though within a smaller space. From Table 8.3 it is also clear that the set of objects in close vicinity to the human agent is an important clue to the human agent’s activity which is captured as being a part of the action space.
Recognizing Human Activities and Actions

Activity recognition using action space as an information channel produces results that are similar to the selected set both at the activity level as well as at the action-level. Note that exploitation of the selected set has been a dominant approach in recognizing human activities (Philipose et al., 2004, Patterson et al., 2005). For certain activities (1, 2, 3, 5, 9), the actions are recognized better using the selected set while there is another subset of activities for which the actions are recognized better using the action space information channel. This is not just a coincidence since activities 1, 2, 3, 5 and 9 are precisely the “prepare” activities, like for instance preparing cake or preparing the table, wherein a set of objects has to be picked up from different places and brought to a particular place. On the other hand, action space is more efficient for activities like having lunch or having breakfast, in which a set of objects is already present at the beginning of the activity and remain present until the end of the activity. The proximity principle discussed earlier is chapter 3 is applicable here where for instance during the action have the dessert, the human agent pushes away the main meal and brings the fruit bowl closer, then pushes the fruit bowl away and brings the coffee jar and cups closer. Refer to Table 8.4 for confusion matrix.

Table 8.3. Precision (P) and recall (R) in percentage (%) for each activity and action using the three information channels (perception space, action space and selected set). The last column represents the precision (P) and the recall (R) obtained by combining the three information channels. The last row represents global values (G) in percentage (%) (Surie et al., 2007b).

<table>
<thead>
<tr>
<th>Activity</th>
<th>Perception space</th>
<th>Action space</th>
<th>Selected set</th>
<th>Combination</th>
</tr>
</thead>
<tbody>
<tr>
<td>#</td>
<td>P (%)</td>
<td>R (%)</td>
<td>P (%)</td>
<td>R (%)</td>
</tr>
<tr>
<td>1</td>
<td>85</td>
<td>100</td>
<td>70</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>80</td>
<td>100</td>
<td>61</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>79</td>
<td>99</td>
<td>43</td>
<td>99</td>
</tr>
<tr>
<td>4</td>
<td>72</td>
<td>100</td>
<td>60</td>
<td>100</td>
</tr>
<tr>
<td>5</td>
<td>80</td>
<td>98</td>
<td>69</td>
<td>98</td>
</tr>
<tr>
<td>6</td>
<td>84</td>
<td>94</td>
<td>67</td>
<td>92</td>
</tr>
<tr>
<td>7</td>
<td>86</td>
<td>99</td>
<td>49</td>
<td>96</td>
</tr>
<tr>
<td>8</td>
<td>81</td>
<td>97</td>
<td>68</td>
<td>95</td>
</tr>
<tr>
<td>9</td>
<td>85</td>
<td>98</td>
<td>80</td>
<td>97</td>
</tr>
<tr>
<td>10</td>
<td>82</td>
<td>99</td>
<td>76</td>
<td>99</td>
</tr>
<tr>
<td>G</td>
<td>81</td>
<td>98</td>
<td>64</td>
<td>98</td>
</tr>
</tbody>
</table>
8.2.4.2 Information Channels Complement Each Other

By combining all three information channels namely perception space, action and selected set, a recognition precision of 89% at activity level and 76% at action level is obtained. Such a high precision at the action level is possible due to the combination of the information channels that represent different and complementary aspects of the human agent activities. In Fig. 8.4, for the activity of doing the dishes, action 1 (clean the dishes) and 2 (dry the dishes on the rack) share similar selected set data. However, the separation is improved by using perception space, which includes the sink for action 1 and the rack for action 2. In some activities, like for instance preparing fried vegetables, the actions take place in a fixed location, and the set of objects remains the same because all the objects needed for this activity are already around the human agent. In such cases, perception space data produces less discriminating information compared to selected set at the action level (see Fig. 8.5). Note that the ground truth was available for all the classifier predictions.
Recognizing Human Activities and Actions

Fig. 8.4. Ground truth vs. activity recognition system I for perception space (left), selected set (center), and combination (right). In this case, perception space is shown to be the dominant information channel for activity recognition (Surie et al., 2007b).

Fig. 8.5. Ground truth vs. activity recognition system I for perception space (left), selected set (center), and combination (right). In this case, selected set is shown to be the dominant information channel for activity recognition (Surie et al., 2007b).
Activity recognition approaches like (Philipose et al., 2004, Patterson et al., 2005) focus only on selected set patterns and encounter difficulties in classifying activities that involve similar sets of objects. For example, (Patterson et al., 2005) keeps track of the number of times an object was touched to differentiate between the activity of preparing the table and the activity of having breakfast that uses a similar set of objects. Such an approach requires many parameters that need to be fine-tuned for individual cases, thereby limiting its scalability to a wider variety of activities.

8.3 Activity Recognition based on Selected and Manipulated Sets

8.3.1 Intra and Extra Manipulation of Everyday Objects

The state changes to everyday objects in the human agent’s environment while performing activities are kept track of by sensing the agent’s selected and manipulated sets. Refer to the Situative Space Model described in chapter 3. The concept of intra manipulation is inspired by the fact that many everyday objects change their internal state when selected and manipulated by a human agent, while the complementary concept of extra manipulation is motivated by the fact that humans commonly arrange and rearrange the spatial relations between everyday objects as part of their activities (Pederson, 2003). For example, a refrigerator could be considered as a container that contains objects like milk packet, juice bottle, cake box, etc. When the human agent removes the milk packet from the refrigerator, then the milk packet is being extra manipulated by virtue of its changed relationship to other objects, specifically the refrigerator. Similarly, a human agent might turn on the stove, which means it has been intra manipulated by virtue of its changed internal state.

The operational definitions of intra manipulation and extra manipulation for activity recognition in the easy ADL ecology are as follows:

- **Intra manipulation (IM).** An operation by a human agent where in an object is selected and manipulated causing a state change to the object’s internal state is known as an intra manipulation. When a human agent interacts with everyday objects, some objects might change their internal state resulting in the following events: objectID {is_grasped, is_released, is_activated, is_deactivated, is_opened, is_closed}. Refer to Fig. 8.6 (right) for the objects that can change their internal state due to a human agent’s interaction with it. The objectID of each object the human agent is holding (the difference between left hand and right hand is not made) between the corresponding is_grasped and is_released events is kept track of. This information is complemented with the objects' internal state information between their corresponding is_activated and
is_deactivated events or is_open and is_closed events (captured every 1 sec) to obtain what is referred to as the intra manipulation information channel. For more information about intra and extra manipulation, refer to (Pederson, 2003).

![Fig. 8.6. Smart objects with volumes sensitive to extra manipulation marked in red color (left) and smart objects that possess internal state change virtual sensors (right) within the easy ADL ecology (Surie et al., 2007a).](image)

- **Extra manipulation (EM).** An operation by a human agent where in an object is selected and manipulated causing an external state change to that object, i.e. a change in the relation of that object to other objects (usually a change in the object’s location) in the environment is known as extra manipulation. When a human agent interacts with everyday objects, some objects might change their external state resulting in the following events: containerID (objectID) [has_entered, has_left]. objectID refers to the object the human agent is currently interacting with, while containerID provides information about the object that contained or will contain the object the human
agent is currently interacting with. Refer to Fig. 8.6 (left) for the objects (marked in red) that can contain other objects including refrigerator, freezer, cupboard, and dining table (all in the virtual reality simulated easy ADL ecology). The external state change information includes whether the object has entered the container or has left the container.

When objects are selected by a human agent, the two events $\text{objectID \{is\_grasped, is\_released\}}$ alone are considered for activity recognition in (Philipose et al., 2004, Patterson et al., 2005). The activity recognition system I described in section 8.2 provided the utility of perception space and action space as novel information channels for activity recognition. In this section, the utility of intra manipulation (selected set + manipulated set causing internal state changes) and extra manipulation (selected set + manipulated set causing external state changes) as information channels to recognize activities with higher precision and to recognize some events more sharply, especially in determining the beginning and end of an activity will be shown. An ambient ecology built to provide reliable assistance to people suffering mild dementia should include accurate activity recognition with sharper event detection as basic requirements.

8.3.2 Activity Recognition System II

As mentioned in section 8.2.2, immersive virtual reality was used in comparing the contributions of the intra manipulation information channel and the extra manipulation information channel for activity recognition. A virtual reality model developed using the Colosseum3D real-time physics platform (Backman, 2005) was used to simulate the wearable sensors and sensors embedded in selected smart objects to capture intra manipulation (IM) and extra manipulation (EM) events. Refer to Fig. 8.7 for the activity recognition system II architecture.

78 object types were used for the evaluation within the easy ADL ecology. Object types include simple object types like fork, knife, plate, etc. that does not change their internal state, complex object types like microwave oven, stove, oven, tap, etc., that can potentially change their internal states, and container object types like fridge, freezer, cupboard, dining table, etc. that can contain other objects. 7 objects have internal states and 11 objects are container objects. We only consider the type of object in recognizing activities, not the identity (e.g. fork_1 and fork_2 are both considered as fork type) similar to activity recognition system I. There are many smart objects with varying capabilities that occur in several activities. For instance, fork, knife, plate, etc., that just communicate their identity are used for several activities like preparing table for lunch, having lunch, having coffee-break, doing the dishes, etc. Similarly, smart objects that communicate their internal state changes like stove, oven, microwave oven, etc., and container objects like fridge, freezer, cupboard, etc., are also used for many activities. This makes the classification problem harder compared to taking an approach where the
Recognizing Human Activities and Actions

217

Activity recognition system is strongly characterized by one or two objects that are unique to the activity. Some activities, like for instance *preparing_rice*, involve objects unique to the activity, in this case the *rice_bag* and the *rice_cooker*. But this does not simplify the classification problem for the following reasons also discussed in section 8.2: 1) the activity recognition system recognizes not only the current activity, but also the current action, and the *rice_bag* and the *rice_cooker* are not involved in all the actions within the activity of *preparing_rice*, only in some; 2) the *rice_bag* manipulation or the *rice_cooker* manipulation might be noise inadvertently created by the human agent while performing another activity; and 3) the recognition system should recognize the activity and the action before they are actually completed to provide appropriate assistance to the human agent, which means that the activity recognition system cannot wait until the unique object is manipulated to recognize the activity and the action.

8.3.2.1 Feature Extraction and Classification

The intra manipulation and extra manipulation information channels consist of sets of events that need to be quantified every second. The quantification scheme used builds $S_A$ and $S_B$ as shown in Fig. 8.7, where $S_A$ represents the set of distinct eventIDs calculated using objectIDs and the object’s internal state, while $S_B$ represents the set of distinct eventIDs calculated using objectIDs, containerIDs and the object’s external state change. The probabilistic generative framework of a hidden-markov model (HMM) is used considering its advantages mentioned in section 8.2. HMMs reduce the system’s configuration space into a finite number of discrete states together with the probabilities for transition between the states. One limitation of HMMs is that the model structure has to be user-defined (also mentioned in section 8.2), which includes the number of states and the connections between the states. The model structure cannot be determined by standard learning.
methods. This should not pose a major problem since the activities recognized are user-defined. The human agent provides ground truth for both activities and actions. Each activity is modeled using a separate HMM with the number of states corresponding to the number of actions within that activity. Similarly, the transitions between states correspond to the transitions between different actions within that activity. Activity recognition system II uses two information channels as shown in Fig. 8.7. Each information channel produces a sequence of observations that are fed into fifteen HMMs (one for each activity). For each information channel, the outputs from the fifteen HMMs are used to build an activity probability vector (\( A_A \) and \( A_B \)) containing the probabilities for each possible activity. The element of the activity probability vector with the highest value gives the human agent’s current activity and its most probable state gives the human agent’s current action.

8.3.2.2 Combining the Information Channels

The two information channels are combined using activity contribution factors (\( W_A \) and \( W_B \)) and action contribution factors (\( w_A \) and \( w_B \)). \( W_A \) and \( W_B \) consist of the recognition precision values for each activity using the intra manipulation information channel and the extra manipulation information channel respectively, while \( w_A \) and \( w_B \) consist of the recognition precision values for each action using the intra manipulation information channel and the extra manipulation information channel respectively. The contribution factors are automatically generated from the training data. The human agent’s current activity is first determined by computing \( R \), the weighted sum of both information channels using the following formula:

\[
R = W_A * A_A + W_B * A_B
\]

where * represents element-by-element multiplication. The element of \( R \) with the highest value gives the human agent’s current activity \( A \). Once the activity is known, the human agent’s current action is determined by calculating \( r \) using the following formula:

\[
r = w_A * pS_A(A) + w_B * pS_B(A)
\]

where \( pS_A(A) \) and \( pS_B(A) \) are the vectors of state probabilities of the HMM representing the human agent’s current activity \( A \) for intra manipulation and extra manipulation respectively. \( w_A \) and \( w_B \) are equal to zero if their respective channel is not supportive to the activity \( A \) determined previously. The element of \( r \) with the highest value gives the human agent’s current action. The two information channels are not combined before classification in order to evaluate their contributions independently and also to combine them based on their contribution in recognizing individual activities. Such an approach also provides the possibility to include additional channels without affecting the overall infrastructure of the activity recognition system II.
8.3.3 Evaluation II

The qualitative evaluation was performed with 5 subjects in the immersive virtual reality easy ADL ecology. 15 activities of daily living were included as shown in Table 8.5.

Table 8.5. List of activities and actions selected based on the AMPS framework (AMPS, 2011).

<table>
<thead>
<tr>
<th>No.</th>
<th>Activity Name</th>
<th>Actions within individual Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Preparing rice</td>
<td>Bring rice bag, Pour rice, Pour water, Add salt, Switch ON rice cooker, Replace rice bag</td>
</tr>
<tr>
<td>2</td>
<td>Preparing vegetables</td>
<td>Bring vegetables, Cut vegetables, Place pan on the stove, Switch ON stove, Fry vegetables in oil, Add spages and salt, Switch OFF stove</td>
</tr>
<tr>
<td>3</td>
<td>Baking cake</td>
<td>Bring baking dish, Switch oven ON, Add eggs, Add milk, Add sugar, Add cake powder, Place baking dish into oven, Switch oven OFF, Remove baking dish</td>
</tr>
<tr>
<td>4</td>
<td>Preparing cake</td>
<td>Get cake, Put cake into microwave oven, Switch ON microwave oven, Remove cake from microwave oven</td>
</tr>
<tr>
<td>5</td>
<td>Preparing coffee</td>
<td>Take coffee powder, Pour water into coffee maker, Switch ON coffee maker</td>
</tr>
<tr>
<td>6</td>
<td>Having a coffee break</td>
<td>Pour coffee, Drink coffee, Cut cake, Eat cake</td>
</tr>
<tr>
<td>7</td>
<td>Having lunch</td>
<td>Have main meal, Have dessert, Have coffee</td>
</tr>
<tr>
<td>8</td>
<td>Preparing table for coffee</td>
<td>Bring cake and coffee, Bring cutlery</td>
</tr>
<tr>
<td>9</td>
<td>Preparing table for lunch</td>
<td>Place mats, Bring cutlery, Bring plates, Bring juice glasses, Bring the food, Place napkins</td>
</tr>
<tr>
<td>10</td>
<td>Doing the dishes</td>
<td>Bring the dishes, Brush the dishes, Rinse the dishes, Wash hands</td>
</tr>
<tr>
<td>11</td>
<td>Preparing apple pie</td>
<td>Bring the baking dish, Bring butter, Bring flour, Bring apples, Cut apples, Switch oven ON, Place baking dish into oven, Switch oven OFF, Remove baking dish</td>
</tr>
<tr>
<td>12</td>
<td>Preparing pasta</td>
<td>Get pasta, Switch stove ON, Fill casserole with water, Add salt, Set the timer and switch the stove ON, Drain with the colander</td>
</tr>
<tr>
<td>13</td>
<td>Preparing pasta sauce</td>
<td>Place pan on the stove, Add oil, Switch stove ON, Bring onions and cut them, Place onions in the pan, Bring tomatoes and cut them, Place tomatoes in the pan, Add spices and salt, Switch stove OFF</td>
</tr>
<tr>
<td>14</td>
<td>Preparing tea</td>
<td>Fill casserole with water, Switch stove ON, Add tea bags, Switch stove OFF</td>
</tr>
<tr>
<td>15</td>
<td>Cleaning the kitchen</td>
<td>Take sponge and spray, Clean the microwave oven, Clean the oven, Clean the stove, Clean the sink, Clean the table, Replace stuff and clean hands</td>
</tr>
</tbody>
</table>

The activities were performed 10 times each by each subject as part of various scenarios. A scenario comprises of a few related activities performed in some sequence. 7 scenarios were used: lunch scenario 1, lunch scenario 2, coffee-break scenario 1, coffee-break scenario 2, baking scenario 1, baking
scenario 2 and cleaning scenario. Some activities were common for several scenarios like the activity of preparing table for lunch which is common to both the lunch scenario 1 and the lunch scenario 2.

Table 8.6. Precision (P) and recall (R) in percentage (%) for each activity and action using the two information channels (intra manipulation and extra manipulation) and by combining the two information channels. The last row represents global values (G) for all activities combined in percentage (%). A# refers to the activity number (Surie et al., 2007a).

<table>
<thead>
<tr>
<th>Activity</th>
<th>Action</th>
<th>Activity</th>
<th>Action</th>
<th>Activity</th>
<th>Action</th>
<th>Activity</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>A#</td>
<td>P (%)</td>
<td>R (%)</td>
<td>P (%)</td>
<td>R (%)</td>
<td>P (%)</td>
<td>R (%)</td>
<td>P (%)</td>
</tr>
<tr>
<td>1</td>
<td>81</td>
<td>100</td>
<td>68</td>
<td>100</td>
<td>90</td>
<td>100</td>
<td>90</td>
</tr>
<tr>
<td>2</td>
<td>47</td>
<td>99</td>
<td>35</td>
<td>99</td>
<td>73</td>
<td>100</td>
<td>73</td>
</tr>
<tr>
<td>3</td>
<td>92</td>
<td>100</td>
<td>76</td>
<td>100</td>
<td>81</td>
<td>94</td>
<td>81</td>
</tr>
<tr>
<td>4</td>
<td>78</td>
<td>98</td>
<td>45</td>
<td>96</td>
<td>45</td>
<td>60</td>
<td>45</td>
</tr>
<tr>
<td>5</td>
<td>89</td>
<td>96</td>
<td>89</td>
<td>96</td>
<td>67</td>
<td>99</td>
<td>67</td>
</tr>
<tr>
<td>6</td>
<td>94</td>
<td>98</td>
<td>91</td>
<td>98</td>
<td>79</td>
<td>99</td>
<td>79</td>
</tr>
<tr>
<td>7</td>
<td>69</td>
<td>100</td>
<td>69</td>
<td>100</td>
<td>94</td>
<td>99</td>
<td>94</td>
</tr>
<tr>
<td>8</td>
<td>57</td>
<td>100</td>
<td>57</td>
<td>100</td>
<td>93</td>
<td>100</td>
<td>93</td>
</tr>
<tr>
<td>9</td>
<td>46</td>
<td>99</td>
<td>46</td>
<td>99</td>
<td>91</td>
<td>100</td>
<td>91</td>
</tr>
<tr>
<td>10</td>
<td>63</td>
<td>98</td>
<td>46</td>
<td>95</td>
<td>98</td>
<td>97</td>
<td>93</td>
</tr>
<tr>
<td>11</td>
<td>63</td>
<td>99</td>
<td>55</td>
<td>99</td>
<td>96</td>
<td>87</td>
<td>96</td>
</tr>
<tr>
<td>12</td>
<td>72</td>
<td>99</td>
<td>44</td>
<td>99</td>
<td>82</td>
<td>100</td>
<td>81</td>
</tr>
<tr>
<td>13</td>
<td>57</td>
<td>98</td>
<td>42</td>
<td>98</td>
<td>90</td>
<td>98</td>
<td>90</td>
</tr>
<tr>
<td>14</td>
<td>93</td>
<td>96</td>
<td>89</td>
<td>95</td>
<td>62</td>
<td>84</td>
<td>62</td>
</tr>
<tr>
<td>15</td>
<td>73</td>
<td>98</td>
<td>26</td>
<td>94</td>
<td>94</td>
<td>99</td>
<td>89</td>
</tr>
<tr>
<td>G</td>
<td>72</td>
<td>98</td>
<td>58</td>
<td>98</td>
<td>82</td>
<td>94</td>
<td>81</td>
</tr>
</tbody>
</table>

All the subjects were allowed to perform the activities in their own way (often in many different ways). When a subject begins performing their activity, each object is in the location where it was last placed in the subject’s previous activity. This makes the experiments realistic compared to having a fixed initial location for individual objects. Cases when the subjects dropped an object on the floor, took the wrong object or performed an inappropriate object state change were also included in our dataset. A real chair was used
Recognizing Human Activities and Actions

for the subjects in the activities of having coffee-break and having lunch that obliged them to sit down. Subjects’ body postures and locomotion within the virtual reality environment were realistic. For instance, the subjects were not allowed to pass through a table, even though it is possible in a virtual reality environment as discussed in section 8.2.

Table 8.7. Confusion matrix where A# refers to the activity number (Surie et al., 2007a).

<table>
<thead>
<tr>
<th>Actual activities</th>
<th>Recognized activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>A# 1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>595</td>
</tr>
<tr>
<td>2</td>
<td>31</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
</tr>
</tbody>
</table>

8.3.3.1 Precision, Recall and Confusion Matrix

The average number of events generated by the intra manipulation information channel is 42 per action, while that of extra manipulation information channel is 3.9 per action. The observation sequence lengths were empirically determined for individual information channels based on a trade-off between the precision and recall values. An optimal observation sequence length of 6 is used for both intra manipulation and extra manipulation information channels. The “Leave-One-Out Cross-Validation” (LOOCV) scheme was used to obtain the precision and recall figures shown in Fig. 8.6. Cross-validation was used to validate the classification considering the limited, but sufficient datasets from the 5 subjects. Refer to Table 8.7 for the confusion matrix.
8.3.3.2 Information Channels Reinforce Each Other at the Activity Level

By combining the two information channels (intra manipulation and extra manipulation), a recognition precision of 92% was obtained at the activity level. Such a high precision is possible due to the combination of the information channels that represent different and complementary aspects of the human agent’s activities. In Fig. 8.8, for the activity of preparing cake, action 2 (Put cake into microwave oven) and action 3 (Switch ON microwave oven) are recognized using intra manipulation as the information channel. This is due to the fact that the internal state of the microwave oven changes from is_closed to is_open and then to is_closed once again during action 2, while the microwave oven changes its internal state from is_deactivated to is_activated and then to is_deactivated during action 3. Action 4 (Remove cake from microwave oven) is confused with action 2 because during both these actions, similar events are generated. In this case extra manipulation provides more reliable information since Microwave Oven (Cake) {has_left} event is generated that is unique for action 4. Action 1 (Get cake) is recognized equally well using both intra manipulation and extra manipulation as information channels. Note that the ground truth was available for all the classifier predictions.

8.3.3.3 Information Channels Complement Each Other at the Action Level

At the action level, extra manipulation shows a good precision, but the recall value of 94% indicates that some actions may not be detected at all times. A closer look reveals that actually 21 actions among the 15 activities are not detected even once, which is not acceptable in building a reliable activity recognition system for ambient ecologies. But by combining intra manipulation and extra manipulation information channels, the recall value is increased from 94% to 99%, thereby allowing all the actions within the 15 activities to be detected at the cost of a lower overall recognition precision.

8.3.3.4 Combining the Information Channels using Weights

At present the combination between intra manipulation and extra manipulation information channels is limited to simple weights. As can be seen from Table 8.6, by combining the information channels the activity recognition precision has improved to 92%. However, the combination at the action level has reduced the recognition precision from 81% (extra manipulation information channel alone) to 70% (combining intra manipulation and extra manipulation information channels). Further research needs to be done in combining the information channels.

The output of the activity recognition system II may sometimes be unstable, especially around the transitions between two activities or between two actions. In order to provide reliable activity support within the easy ADL ecology, the output of the activity recognition system II must be smoothened.
which introduces delays between the actual starting of an activity or action
and the moment when the system makes a guess about the human agent’s
current activity or action. An activity is recognized on average after 30 events.
At the action level, this delay varies between 7 and 24 events (13 on average),
which means that in some cases (especially if the duration of an action is
short) the system may guess the human agent’s current action with a delay of
one or two actions. According to (Ward et al., 2006), it is important to not only
know the recognition accuracy, but also the timing of the recognition while
performing the activities to enable its feasibility for specific applications.

![Fig. 8.8. Intra manipulation and extra manipulation information channels complement each other for activity recognition (Surie et al., 2007a).](image)

### 8.3.3.5 Delay in Recognizing Activities and Actions

The output of the activity recognition system II may sometimes be unstable,
especially around the transitions between two activities or between two
actions. In order to provide reliable activity support within the easy ADL
ecology, the output of the activity recognition system II must be smoothened
which introduces delays between the actual starting of an activity or action
and the moment when the system makes a guess about the human agent’s
current activity or action. An activity is recognized on average after 30 events.
At the action level, this delay varies between 7 and 24 events (13 on average),
which means that in some cases (especially if the duration of an action is
short) the system may guess the human agent’s current action with a delay of
one or two actions. According to (Ward et al., 2006), it is important to not only

know the recognition accuracy, but also the timing of the recognition while performing the activities to enable its feasibility for specific applications.

8.3.3.6 Implications for Activity Support within the easy ADL Ecology

Activity recognition system I (refer to 8.2) using different information channels yielded a precision of 89% among 10 activities of daily living. A major reason for attempting to improve on this by exploring complementary information channels is to recognize activities with higher precision (activity recognition system II yielded a recognition precision of 92% among 15 activities of daily living) and obtain sharper events that can be used for both activity recognition, and for providing real-time activity-centered support within the easy ADL ecology. Examples of sharper events include turning ON a stove or taking out juice packet from the fridge. Events that are generated using extra manipulation and intra manipulation information channels are sharper compared to the bare objects grasped or released events, since extra manipulation and intra manipulation information channels contain information about specific objects' state change. Events generated using the perception space and the action space information channels (refer to section 8.2) do not yield sharp events that can directly be used by the ambient intelligence applications and other components of the personal activity-centric middleware in providing assistance to the human agent within the easy ADL ecology. Even though sharper events were obtained using extra manipulation and intra manipulation information channels, there remains some uncertainty and delay in the recognition. To address this limitation, which is inherent to the probabilistic nature of activity recognition systems in general, data mining on the training data was made in order to extract events that always recur for a particular activity. Such events are referred to as mandatory events. For instance, if a human agent switches ON the stove during the activity of preparing vegetables in all the training data, then switching ON the stove is considered to be a mandatory event. Such mandatory events along with the human agent's current activity and action are used by the personal activity-centric middleware and the ambient intelligence application in providing assistance to the human agent.

8.3.3.7 Improving the Granularity of Intra Manipulation Information Channel

The internal state changes of objects represent information important to capture in recognizing human activities and actions. Evaluation II presented in this section included only a selected number of smart objects that communicate their state change information to avoid too much sensing and computation going on in the smart objects (simulated in the virtual reality setup). Among those selected smart objects, objects that are unique for individual activities like rice cooker, coffee maker, etc. have contributed well to activity recognition. However, devices like stove and oven that are common
for several activities provide less information to the activity classifier and introduce noise in the activity recognition. This is because the internal state of the stove or the oven is sensed with coarse granularity (only on/off states). By improving the sensing of the smart objects’ internal states, using finer granularity of states, like for instance, sensing the temperature of the stove, there is a reasonable chance of improving the recognition at both the activity and the action level. The temperature of the stove changes differently during the activity of preparing tea compared to the activity of preparing pasta sauce. Such fine-grained sensing of the internal states of smart objects in a living laboratory setup is described in chapter 6; however the data is yet to be used for activity recognition.

8.3.3.8 Exploring the Spatial Relations between Objects

At present, the extra manipulation information only includes the simple relationship between the object the human agent is currently interacting with and any container changes due to such interaction. However, there may be some relationship between the object the human agent is currently interacting with and the other objects that are inside or on the container object. For instance, on a dining table there may be table mats and also plates, knives and forks during the activity of having lunch. By removing the table mats, not only the relationship they had with the dining table changes, but also their relationship with plates, knives and forks that are on the dining table. Such relations among objects and their spatial relationships will be explored in the future. Also, note that the version of the situative space model that is described in chapter 3 does not include spatial relationships among objects, but there is potential for future extensions of the model in this regard.

8.4 Discussion

8.4.1 Transferability to Real-World Applications

Immersive virtual reality simulations offer “clean” sensing with little noise and uncertainty in the collected signals, as opposed to sensing in the real, physical world. Selected set and action space require the identification of objects close to the human agent’s body which can be reliably obtained using RFID technology (Finkenzeller, 2003) as shown in many applications including (Philipose et al., 2004, Patterson et al., 2005, Pederson, 2001). Action space would require an RFID reader with higher power compared to the selected set due to the larger reading range necessary. Sensing perception space could be more complicated than selected set and action space due to range limitations with passive RFID technology. Capturing the extra manipulation information introduces many challenges including the difficulties in limiting the volume of the container object that is sensitive to extra manipulation events and in attaching RFID readers on devices like oven
that might be used at high temperatures. Some passive tags can handle high temperatures, and future investigations would include exploring RFID readers that can handle high temperatures. Similarly, there are issues that need to be solved in capturing intra-manipulation information. Refer to chapter 6 and 7 for further information about sensing the state changes to objects and tracking the situative spaces within a living laboratory easy ADL ecology, which has yielded reasonable results.

The focus of this chapter is not on the technology that exists today, but instead to investigate the contributions the different situative spaces and sets can make in recognizing everyday human activities and actions, with the assumption that sufficient technology to capture this information with better accuracy will be available in the near future. Even though external validity (Mitchell and Jolley, 2001) cannot be guaranteed, the approach taken is a novel one and is primarily intended for guiding the development of the activity recognizer, an important component of the personal activity-centric middleware described in chapter 4.

8.4.2 Scalability to Number of Activities

Human agents may require support for a potentially large number of activities within the easy ADL ecology. Since individual models are used for each activity, the approach proposed in this chapter for activity recognition should be able to cope with the increase in the number of activities. But by increasing the number of activities, the recognition accuracy might decrease. The 10 activities used for evaluating the activity recognition system I and the 15 activities used for evaluating the activity recognition system II took place mainly in a kitchen and dining hall setup. One could estimate that the total number of activities of the same kind, in the same kind of rooms, and at the same level of granularity, is less than 10 times as many. It is likely that this approach would be able to handle the scaling factor. Furthermore, one could believe that although the total repertoire of everyday activities for a person may be large, they may be fairly well distributed over different rooms in a home environment. This would mean a moderate number of possible activities for each room that need to be distinguished between.

8.4.3 Adaptation to Variations in Activity Patterns

Human agents may perform the same activity in several different ways depending on the situation and other factors. This issue was addressed to some extent by including such variations in the training data. However, with time, human agents may change the underlying structure of some activities. This is potentially possible when for instance, a human agent’s habits change over time or when new objects are introduced that change the way certain activities are supposed to be performed. Thus, a modeling approach capable of dynamically changing its internal model with time would be desirable, but represents a major challenge. In the future, it is important to investigate the
possible techniques to make the HMMs adaptive in the long term as well as the short term.

8.4.4 Online and Unsupervised Training

The amount of time and efforts spent by the subjects in the immersive virtual reality setup for training the activity recognition systems was a lot. One of the main reasons for it was that the training was performed offline, i.e. data was collected during a separate training session where the subjects performed individual activities and provided the ground truth. Online training enables a human agent to perform activities without an explicit training phase and the activity recognition system gets trained during the actual performance of activities. While online training approach could reduce the performance of the activity recognition system, it should definitely be a future work to reduce the burden on the subjects. Similarly, unsupervised training is another way forward for the future. While unsupervised training might make it had for an activity recognition system to model complex activities, it will mitigate the efforts required from the subjects. After all, activity recognition systems are built to improve a human agent’s experience within an ambient ecology and not to frustrate or burden them with mundane training activity.

8.4.5 Handling Interrupted and Interleaved Activities

Human activities may be interleaved with another activity, or interrupted often which is to some extent addressed in (Patterson et al., 2005). While it is important to include features in an activity recognition system that handles interleaved and interrupted activities, knowing exactly when a human agent has shifted from one activity to another is difficult to determine in real-time. By introducing “key” object states or state changes that are associated to individual activities, the challenge of knowing when an activity has shifted course could be recognized.

8.4.6 Recognizing Activities performed by Multiple Human Agents

While the activity recognition problem is simplified by taking single human agent in an ambient ecology scenario, the complexity shoots-up instantly when multiple human agents perform activities in an ambient ecology. Further discussion on this topic is presented in chapter 9.
Chapter 9

Discussion and Conclusion

This chapter presents a summary of the research presented in this thesis and discusses the contributions. The areas that merit future research are also presented.

9.1 Summary

In chapter 1, an introduction to Ambient Intelligence (AmI) as an emerging computing paradigm that envisions computing to be a part of everyday environments centered on human agents is presented. We briefly tried to understand the evolution of computing, the present state of everyday artifacts and the possibilities of integrating them in moving towards the visions of ambient intelligence.

In chapter 2, a summary of the background and related work in the field of ambient intelligence with a focus on human-computer interaction was presented. Several problems that were not addressed by existing interaction paradigms and approaches were identified:

5. Lack of support for human-environment interaction: Most of the existing work focuses on human interaction with a single device and is typically device-centric.
6. Lack of support for human-centered interaction: Most of the existing work focuses on specific human-related context ignoring their body, situation and current activities as a whole in facilitating interaction.
7. Support restricted to limited input and output modalities: Most of the existing work ignores the possibilities of viewing a human agent’s perceptual and action capabilities as a whole that spans across multiple modalities at varying levels of granularity for presenting and obtaining information.
8. Lack of support for uniformly handling both physical and virtual aspects in facilitating interaction: Most of the existing work does not treat artifacts, situations and activities that spans across the physical and the virtual realm alike.

In chapter 3, the above mentioned problems were taken as a starting point in exploring and describing a novel interaction paradigm centered on human agents, referred to as egocentric interaction. Also,

1. Cognitive science theories like situated action, embodied cognition, activity theory, in conjunction with human-related factors like perception and action, attention, and intention were surveyed for inspiration in developing egocentric interaction concepts. This
approach is in contrast to related interaction paradigms that are inspired by technological advancements.

2. Principles like situatedness and the proximity principle were used in developing the Situative Space Model useful for: (a) describing a human agent’s situation; (b) modeling and recognizing their activities; and (c) facilitating situated interaction based on their perception and action capabilities.

3. The physical-virtual equity principle was useful in treating the physical and the virtual realms alike and in integrating them creating: (a) physical-virtual artefacts; (b) physical-virtual situations; and (c) physical-virtual activities.

4. The novel approach of replacing traditional concepts like input and output with human “perception” and “action” facilitates: (a) interaction through multiple modalities determined by human capabilities; and (b) interaction with multiple devices once again determined by human capabilities.

In chapter 4, the transition from viewing egocentric interaction as a conceptual framework to the actual application of it in designing an ambient ecology was made. The easy ADL ecology was presented as an infrastructure for ambient intelligence. Also,

1. Important constituents of an ambient ecology namely: smart objects, a personal activity-centric middleware, ambient intelligence applications and human agent(s) were identified.

2. Several questions related to smart objects like: (a) what are smart objects? (b) how to model and represent smart objects? etc. were addressed. Illustrations of smart objects developed for the easy ADL ecology were presented.

3. Important features and functionalities of a middleware developed for ambient ecologies and based on egocentric interaction were identified: (a) to manage smart objects; (b) to monitor human situations; (c) to infer their activities and actions; and (d) to manage their interaction with ambient intelligence applications.

4. Important attributes for virtual objects part of ambient intelligence applications were identified. Such virtual object attributes in conjunction with information generated by the middleware about a human agent, their situational and activity context, their environmental context, etc. were used in defining interaction management rules useful in facilitating human interaction with virtual objects.

In Chapter 5, the description of a user-experience evaluation of the “proof-of-concept” easy ADL ecology was presented. As an initial step, a set of user experience factors were identified that shed light on the concepts described as egocentric interaction and its usefulness in designing ambient ecologies. This 20-subject-study brought to surface the non-ergonomic and prototypical nature of the “proof-of-concept” easy ADL ecology (e.g. large amounts of cables on the floor and walls, obtrusive wearable components) and privacy concerns in social context. However, a clear majority of the subjects were very positive
towards using a future more ergonomic and less obtrusive version of the easy ADL ecology, as they regarded the services offered (e.g. situation dependent and multimodal access to virtual objects) as useful. While it is impossible to say how a corresponding ambient ecology would have scored if not being designed according to egocentric interaction principles, it is clear that several of the most praised features are strongly associated with those principles.

Chapters 6, 7 and 8 address the technical and technological challenges involved in implementing the easy ADL ecology. In particular, the focus is on the challenges addressed by the object manager (chapter 6), the situation monitor (chapter 7) and the activity recognizer (chapter 8). In chapter 6, tracking the state changes to physical and virtual objects in the easy ADL ecology was identified as an important challenge. While there are several approaches to tracking the state changes, the direction taken was:

1. Investigation, implementation, deployment and evaluation of a simple ZigBee-based wireless sensor networking of smart objects to track the state changes to physical objects.
2. Implementation and evaluation of simple hand gesture and speech recognition engines useful in keeping track of virtual object manipulations.

In chapter 7, tracking the situative spaces in terms of physical and virtual objects was identified as an important challenge. Also,

1. Different modalities through which objects are perceived and acted upon by human agents were identified.
2. Approaches to track objects in those different modalities were investigated.
3. WLAN signal strength-based situative space tracking was selected as a primary approach for tracking the situative spaces along several modalities. Implementation, deployment and evaluation of this proximity based object tracking system were presented.

In chapter 8, modeling and recognizing human activities was identified as an important challenge. The feasibility of the different information channels generated by the situative space model for recognizing human activities and actions was investigated. In particular,

1. Perception space, action space and selected set were investigated as information channels in activity recognition system I. The implementation details and the evaluation results in an immersive virtual reality simulated easy ADL ecology were reported.
2. Selected set and manipulated set (formed by both intra manipulation and extra manipulation) were investigated as information channels in activity recognition system II. The implementation details and the evaluation results in the immersive virtual reality simulated easy ADL ecology were reported.


9.2 Contributions

To summarize the contributions of this thesis,

1. A novel interaction paradigm referred to as egocentric interaction is developed that supports: (a) human-environment interaction as opposed to human interaction with individual devices; (b) human-centered interaction by unifying human body, their situations and activities as important contextual information; (c) human “perception” and “action” instead of input and output thereby expanding the multimodal design space based on human capabilities; and (d) physical-virtual equity allowing human agents to exist as dual citizens in both the physical and the virtual world.

2. Implementation of the easy ADL ecology to validate the concepts of egocentric interaction as a design tool and to illustrate the concepts more concretely. In particular, smart object systems with varying capabilities are developed in conjunction with a personal activity-centric middleware for handling important challenges within ambient intelligence namely: (a) managing smart objects; (b) tracking a human agent’s situation; (c) recognizing their activities and actions; and (d) managing their interaction with ambient intelligence applications. This eases the development of ambient intelligence applications where the focus can be on the domain knowledge instead of the challenges inherent in an ambient ecology.

3. Development of living laboratory and immersive virtual reality simulated smart environments useful as a test bed for further research in ambient intelligence.

9.3 Limitations and Future Work

The research field of ambient intelligence has a large scope and in this thesis we have not addressed all the challenges and issues. The following are some of the challenges that are worth exploring in the future.

9.3.1 Support for Multiple Human Agents in an Environment

One of the major simplifications made in this thesis is to target the research work (both the conceptual and the implementation parts) for a single human agent in an ambient ecology. While this is not a problem in itself especially during the initial research phase where simplification means the possibility for clearer investigation of the concepts and the prototypes, long-term research in ambient intelligence obliges the need to take multiple human
agent scenarios in an ambient ecology. Such a view affects the modeling and the system development in many ways:

- **Situative space model**: the model described in this thesis is centered on a single human agent especially with a focus on their bodily situation in an environment. This removes the social modeling challenges like the situative spaces that are shared by a human agent with other human agents. The current situative space model should be extended to incorporate: (1) other human agents within a human agent’s situative spaces; (2) the spaces that are shared with other human agents; (3) the objects that are shared with other human agents; (4) introduce the dimension of time useful in situations when objects are shared by time-multiplexing; and (5) introduce space-multiplexing for sharing the situative spaces among multiple human agents.

- **Recognizing human activities and actions**: the activity recognition systems built are intended for single human agent in an ambient ecology. They are to address challenges like: (1) recognizing parallel activities; (2) activities that get interrupted and resumed often; and (3) activities that can be performed in many different ways in the future. Introducing multiple human agents add complications in recognizing activities. It is not only important to know the state changes to objects in recognizing activities but also to know who made those state changes, if those state changes are part of a private activity, collaborative activity, or conflicting activity. Issues like sharing the rights to perceive and manipulate objects surface depending on the type of activities (private, collaborative, conflicting, etc.) and other contextual conditions which are to be addressed in the future.

- **Design complexity of smart objects**: the smart objects presented in this thesis are intended to support a single human agent in an ambient ecology. When such smart objects are to be shared among human agents in the same ambient ecology, the design complexity of smart objects increases. Typical example includes two human agents trying to access the wall display wherein the wall display should include mechanisms to handle conflicts. Issues like when to provide access to a specific human agent, how will its resources be shared, is privacy and trust of all human agents part of the ambient ecology maintained, etc. pops-up and mechanisms to handle such challenges increases the complexity of the smart objects.

### 9.3.2 In-Situ Evaluation of the easy ADL Ecology

Another limitation of the work presented in this thesis is the lack of in-situ evaluation, i.e. evaluation in the real-world context instead of a living laboratory setup or an immersive virtual reality simulation. In-situ evaluations usually hold a higher external validity in comparison to results obtained in a laboratory setup. However, the approach taken of using
immersive virtual reality as an initial development environment followed by physical living laboratory setup seems practical to reduce the development cost in moving towards a functional, robust and stable prototype. The following future works are required in order to perform in-situ evaluations of the easy ADL ecology:

- **Design of smart objects**: the smart objects designed for the “proof-of-concept” easy ADL ecology is not suitable for everyday use in the long run. For instance, the electronics is not properly shielded to conditions like washing and cleaning, rough everyday usage, etc. The ergonomics and usability of many of the smart objects are still a question to be answered: too many wires and cables, usage of inappropriate design material, limited focus on aesthetics and style, no power harvesting, etc. are issues to be addressed.

- **Shorten the training phase**: some of the personal activity-centric middleware components like the activity recognizer and the situation monitor involve extensive training periods making it infeasible and unpractical for real-world context. Establishing the situative space thresholds and building activity models should be performed online, i.e. removing the need for an explicit training phase and integrating the tracking or the recognition phase with the training phase with appropriate feedback from the human agent to the recognition engine.

- **Building functional ambient intelligence applications**: the ambient intelligence applications developed for the prototypical easy ADL ecology are either “mock-up” applications or applications with limited functionalities. Even though such applications are sufficient to explore and understand concepts, in-situ evaluations demand ambient intelligence applications with domain knowledge and provide functionalities that are useful and/or critical in real-world context. At the end of the day, the infrastructure and the conceptual framework proposed are developed to facilitating human interaction with ambient intelligence applications where applications with limited functionalities would limited a human agent’s experience of the easy ADL ecology.

- **Integration of the personal activity-centric middleware components**: The personal activity-centric middleware is functional, but its components are built using different techniques and technologies, developed using different programming languages, and evaluated in different environments. There is a need to standardize the individual components and to integrate it into a single middleware with clearer API that makes it easier for the application developers to build ambient intelligence applications.
9.3.3 Improving the Environment for User Studies

This thesis has presented several user studies for evaluating both the technical aspects of the easy ADL ecology, as well as the qualitative user-experience aspects. Both the immersive virtual-reality simulation environment and the living laboratory environments lacked standardized techniques and technologies useful for capturing user data, efficiently categorize them useful for further analysis, and at the same time address sensitive issues like user privacy, personal data, etc. in a unified manner. Due to the lack of time and resources, many of the studies presented lacked a systematic approach with proper infrastructure for analyzing user data. Separate software was built for collecting activity-related data, situative spaces-related data, and wireless sensor network data. Hand-written notes, and ad-hoc usage of camera filming were used which should be improved in the future to make it easier for extracting results from the user studies.

9.4 Conclusion

This thesis has explored and presented a novel interaction paradigm named as egocentric interaction for modeling and facilitating human-environment interaction within ambient ecologies. This human-centered interaction paradigm offers computing support for everyday human activities that spans the physical-virtual gap by taking the human agent’s body, situation and activities as important contextual information. A situative space model and the physical-virtual design perspective are cornerstones for egocentric interaction.

Typical challenges in the field of ambient intelligence like activity and action recognition, situation and object tracking, smart object management and interaction management have been addressed using a personal activity-centric middleware targeting egocentric interaction, thereby easing the development of ambient intelligence applications and ambient ecologies. The prototypical easy ADL ecology (both the immersive virtual reality simulation and the living laboratory environments) have served as indicators for how egocentric interaction can be used both as a tool for design and for analysis of ambient intelligence systems. The positive results obtained from the user-experience evaluation of the easy ADL ecology marks a promising start, the “secret” for future successful explorations of egocentric interaction as an interaction paradigm for ambient intelligence.
References


ALVAREZ, A., ARCOS, A. & HERRASTI, N. 2006. AMEC - facilitating a user centred evolution to ambient intelligent environments. 2nd IET International Conference on Intelligent Environments.


Discussion and Conclusion


Madison, Wisconsin, United States: American Association for Artificial Intelligence.


Discussion and Conclusion


Discussion and Conclusion

the Move to Meaningful Internet Systems. Springer Berlin / Heidelberg.


JAMES, W. 1890. The principles of psychology, New York, Dover Publications.


KATTERFELDT, E.-S., DITTERT, N. & SCHHELHAW, H. 2009. EduWear: smart textiles as ways of relating computing technology to everyday


Discussion and Conclusion


SIGCHI conference on Human factors in computing systems. Ft. Lauderdale, Florida, USA: ACM.


MÄKELÄ, A. & SURI, J. F. Supporting Users' Creativity: Design to Induce Pleasurable Experiences. International Conference on Affective
Discussion and Conclusion


SJÖLIE, D. 2007. Simulation and distribution of data for the easyADL project. HPC2N, Umeå University.


